Mapping and Data Collection Process Documentation

Detailed GIS and mapping methods used for the SHAPE project

About the Study of Housing and Population Erosion (SHAPE)

SHAPE is a pilot project designed to ensure rural communities are fully seen and counted. Using McDowell County, West Virginia as a starting point, SHAPE addresses a critical challenge: housing units that are missing or inaccurately represented in official address systems due to economic decline, natural disasters, and aging infrastructure.

By combining GIS mapping with on-the-ground verification, SHAPE identifies where people live today, especially in areas that are often overlooked. Improving address accuracy strengthens emergency response, community planning, and future U.S. Census Bureau (Census Bureau) efforts. Just as importantly, SHAPE is intentionally designed to be shared. The tools, lessons learned, and processes developed through this pilot can be adapted by other rural and disaster-impacted communities to support accurate counting, resource access, and long-term preparedness. SHAPE efforts improved on existing McDowell County Census estimates by 45%.

About West Virginia University's Health Affairs Institute (HAI)

HAI at West Virginia University partners with communities, organizations, and government agencies to turn data into action. Our work focuses on helping communities understand challenges, plan effectively, and access the resources they need to thrive.

HAI brings together research, data analysis, project management, and training to support practical, real-world solutions, especially in rural and underserved areas. HAI works alongside local partners to design tools, strengthen systems, and support long-term recovery and growth.

Through projects like SHAPE, the Institute helps ensure communities are visible, counted, and prepared by improving data accuracy, supporting informed decision-making, and sharing models that others can use when facing similar challenges.

Find us: https://healthaffairsinstitute.org/

About Local Update of Census Addresses (LUCA)

LUCA program, established by the Census Address List Improvement Act of 1994, provides tribal, state, and local governments with the unique opportunity to review and improve the Census Bureau's master address list. This collaborative process ensures accurate geographic coverage, which directly impacts the fair distribution of federal funds and political representation. Without correct addresses, communities may miss out on funding and governmental representation.

Participation in LUCA is voluntary and occurs several years prior to the census. Ahead of the 2030 Census, governments will receive LUCA invitations and information beginning May 2027, with data submissions due March 2028.

Please access the website for more information and dates: https://www.census.gov/programs-surveys/decennial-census/about/luca.html

About Robert Wood Johnson Foundation (RWJF)

RWJF is a leading national philanthropy paving the way together toward a further where health is no longer a privilege, but a right. What started as a modest mission to support vulnerable populations in local communities has transformed into a commitment to address structural racism, a root cause of health inequities in the United States. Our understanding and our commitment have deepened, and we continue to push ourselves to be more transparent and accountable as we work to create a more equitable society, centered in health and well-being.

Learn more at: Robert Wood Johnson Foundation | RWJF

Table of Contents

Table of Contents 4
List of Tables 6
List of Figures 6
Document Acronyms 7
1 Mapping Overview 8
1.1 Purpose 8
1.2 Scope 8
2 Data Sources 8
2.1 Geographic Datasets 8
2.2 Non-Spatial Data 20
3 Software and Tools 20
3.1 GIS Software 20
3.2 Supporting Software 20
4 Data Preparation 20
4.1 Data Quality Review 20
4.2 Coordinate System Standardization 20
4.3 Data Cleaning 21
4.4 Data Integration 21
5 Spatial Analysis Methods 22
5.1 Study Area Definition 22
5.2 Geoprocessing procedures, parameters and, methods 22
6 Quality assurance and validation 36
6.1 Accuracy assessment 36
6.2 Comparison with reference datasets 37
6.3 Error rates 37
6.4 Peer review / QA procedures 37
6.5 Limitations 38
7 Mapping and visualizations 39
7.1 Cartographic standards 39
7.2 Map production 39
8 Outputs and Deliverables 39
8.1 GIS Data Products 39

8.2 Maps 39
9 Results 40
9.1 Summary Statistics 40
9.2 Analytical findings 42
10 References 50
Appendix A: GIS Workflow Diagram 52

List of Tables

Table 1: All geographic data utilized in this project applicable to McDowell County, West Virginia 10
Table 2: Geocoding results 21
Table 3: Residential point dataset 24
Table 4: Web maps created for the SHAPE project 40
Table 5: Summary Statistics 41

List of Figures

Figure 1: Geocoding results 22
Figure 2: Example of routes with assigned stops and census tract boundaries 31
Figure 3: Example of routes with assigned stops and census tract boundaries 32
Figure 4: Example of web map 33
Figure 5: Example of web map 34
Figure 6: Mountain State Assessment of Trends in Community Health (MATCH) undeliverable map 42
Figure 7: Original GIS file address density map 43
Figure 8: ACS occupied address estimation map 44
Figure 9: Original GIS address counts map 45
Figure 10: Field collected address counts map 46
Figure 11: Address counts percent change map 47
Figure 12: Bivariate choropleth comparison map 48
Figure 13: ACS estimation compared to field counts 49

Document Acronyms

The following acronyms are used throughout this document:

Acronym	Definition
ACS	American Community Survey
GIS	Geographic Information Systems
HAI	West Virginia University Health Affairs Institute
LUCA	Local Update of Census Addresses
MATCH	Mountain State Assessment of Trends in Community Health
SHAPE	Study of Housing and Population Erosion
Census Bureau	U.S. Census Bureau
RMSE	Root Mean Square Error

1 Mapping Overview

1.1 Purpose

The purpose of the Study of Housing and Population Erosion (SHAPE) project is to document and share all processes and methods applied during the project lifecycle. The main objective of this project is to gather updated residential address data in time for the Local Update of Census Addresses (LUCA) prior to the 2023 Census, and McDowell County was selected as a pilot project due to its declining population and its geographic isolation.

To validate and collect updated address data in the county, basic Geographic Information Systems (GIS) methods and applications were applied with the guidance and support of GIS professionals from various organizations, such as West Virginia Department of Transportation, the McDowell County GIS Specialist, and the Census Bureau.

The purpose of this document is to share detailed information about the data and methods used to validate addresses in McDowell County, West Virginia. This document describes data sources, methods, and analyses applied to collected data in detail. All data sources are listed and described, and are accessible as part of the toolkit. Methodology and independent analyses steps are discussed in chronological order, from initial data cleaning to variance mapping.

You can view SHAPE's Address Validation Toolkit by visiting healthaffairsinstitute.org/featured-work/shape.

1.2 Scope

Updated and field-validated address data will be shared with the Census Bureau, McDowell County GIS Department, and the Robert Wood Johnson Foundation as part of the project toolkit.

2 Data Sources

2.1 Geographic Datasets

SHAPE leadership developed meaningful working relationships with key stakeholders prior to project initiation. Sharing complex GIS data requires a certain level of trust and close cooperation between project participants and stakeholders. It is not uncommon in the GIS realm to experience technical difficulties with collecting and validating GIS data, and mutual support ensures a solid working foundation required for reliable and accurate results.

Numerous GIS datasets were received from key project stakeholders and utilized in this project. Other geographic datasets were developed by SHAPE staff to aid field data collection and all geographic files utilized are summarized in Table 1. All geographic data utilized in this project is applicable to McDowell County, West Virginia.

Table 1: All geographic data utilized in this project applicable to McDowell County, WV

Table 1: All geographic data utilized in this project applicable to McDowell County, West Virginia

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
All_Routes_McDowell_County	McDowell County GIS Department (GIS Specialist)	Shapefile / Feature Class	GCS NAD 1983	NA	All_Routes_McDowell_County.zip
McDowell_Sites	McDowell County GIS Department (GIS Specialist)	Layer file	GCS NAD 1983	A .lyrx file is a visualization file created from a shapefile for sharing purposes to prevent changes to the original vector data.	McDowell_Sites_2025.lyrx
McDowell_Sites_Export	HAI analyst	Shapefile / Feature Class	GCS NAD 1983	Export .lyrx file to a .shp to create a spatial vector file. This file serves as the baseline for variance mapping.	McDowell_Sites_Export.zip
partnership_shapefiles_21v2_54047	West Virginia Partnership Shapefile Batch Download; Census Bureau Boundary and Annexation files	Shapefile / Feature Class	GCS NAD 1983	NA, 9 total census tracts in McDowell County	partnership_shapefiles_21v2_54047.zip
mcdowell_R1-R6	McDowell County GIS Department (GIS Specialist);	Shapefile / Feature Class /	WGS 1984	Select by attribute from 'PointType AddrClass' in McDowell_Sites_Export shapefile: R1 through R6 which are	mcdowell_R1-R6.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
	analysis by HAI analyst	Geodata base		residential points. Geocoded all residential points to create XY Coordinates for Route Planning. Transformed projection to WGS 1984 for AGOL Route Plan analysis.	
censustract_9536	United States Census Bureau	Shapefile / Feature Class	WGS 1984	Using partnership_shapefiles_21v2_54047, select by attribute each specific census tract polygon to create separate census tract polygon for use in Route Planning analysis.	census_tract_9536.shp.zip
censustract_9536_residences	McDowell County GIS	Shapefile / Feature Class	WGS 1984	Using mcdowell_R1-R6, select by location within census_tract_xxxx polygon, those residential points entirely within the census tract polygon to create separate point shapefile of residences within census tract. A total of 656 residential points in this tract.	censustract_9536_residences.zip
assigned_stops_ct9536	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step above. Assigned stops are the residential point locations along each route within each census tract. A total of	assigned_stops_ct9536.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				846 residential units were verified or added in this census tract.	
routes_ct9536	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step. Routes are polylines along ESRI's primary and secondary roads in McDowell County within each census tract. Five routes were generated for this census tract.	routes_ct9536.zip
ct9536_new_addresses	HAI field workers	Shapefile / Feature Class	WGS 1984	New address points were separated from the assigned_stops_ct95xx feature class using the select by attributes function. New points were exported and reverse-geocoded to create XY coordinates and to verify street address data collected in the field. A total of 83 new addresses were collected in this census tract.	ct9536_new_addresses.zip
censustract_9538	United States Census Bureau	Shapefile / Feature Class	WGS 1984	Using partnership_shapefiles_21v2_54047, select by attribute each specific census tract polygon to create separate census tract	census_tract_9538.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				polygon for use in Route Planning analysis.	
censustract_9538_residences	McDowell County GIS	Shapefile / Feature Class	WGS 1984	Using mcdowell_R1-R6, select by location within census_tract_xxxx polygon, those residential points entirely within the census tract polygon to create separate point shapefile of residences within census tract. A total of 1,500 residential points in this census tract.	censustract_9538_residences.zip
assigned_stops_ct9538	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step above. Assigned stops are the residential point locations along each route within each census tract. A total of 1,608 residential units were verified or added in this census tract.	assigned_stops_ct9538.zip
routes_ct9538	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step. Routes are polylines along ESRI's primary and secondary roads in McDowell County within each census tract.	routes_ct9538.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				Eight routes were generated for this census tract.	
ct9538_new_addresses	HAI field workers	Shapefile / Feature Class	WGS 1984	New address points were separated from the assigned_stops_ct95xx feature class using the select by attributes function. New points were exported and reverse-geocoded to create XY coordinates and to verify street address data collected in the field. A total of 91 new addresses were collected in this census tract.	ct9538_new_addresses.zip
censustract_9539	United States Census Bureau	Shapefile / Feature Class	WGS 1984	Using partnership_shapefiles_21v2_54047, select by attribute each specific census tract polygon to create separate census tract polygon for use in Route Planning analysis.	census_tract_9539.zip
censustract_9539_residences	McDowell County GIS	Shapefile / Feature Class	WGS 1984	Using mcdowell_R1-R6, select by location within census_tract_xxxx polygon, those residential points entirely within the census tract polygon to create separate point shapefile of residences within census tract. A total of 1,077	censustract_9539_residences.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				residential units in this census tract.	
assigned_stops_ct9539	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step above. Assigned stops are the residential point locations along each route within each census tract. A total of 1,193 residential units were verified or added in this census tract.	assigned_stops_ct9539.zip
routes_ct9539	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step. Routes are polylines along ESRI's primary and secondary roads in McDowell County within each census tract. Six routes were generated for this census tract.	routes_ct9539.zip
ct9539_new_addresses	HAI field workers	Shapefile / Feature Class	WGS 1984	New address points were separated from the assigned_stops_ct95xx feature class using the select by attributes function. New points were exported and reverse-geocoded to create XY coordinates and to	ct9539_new_addresses.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				verify street address data collected in the field. A total of 87 new addresses were collected in this census tract.	
censustract_9540	United States Census Bureau	Shapefile / Feature Class	WGS 1984	Using partnership_shapefiles_21v2_54047, select by attribute each specific census tract polygon to create separate census tract polygon for use in Route Planning analysis.	census_tract_9540.zip
censustract_9540_residences	McDowell County GIS	Shapefile / Feature Class	WGS 1984	Using mcdowell_R1-R6, select by location within census_tract_xxxx polygon, those residential points entirely within the census tract polygon to create separate point shapefile of residences within census tract. A total of 1,400 residential units in this census tract.	censustract_9540_residences.zip
assigned_stops_ct9540	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step above. Assigned stops are the residential point locations along each route within each census tract. A total of 1,532 residential units were	assigned_stops_ct9540.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				verified or added in this census tract.	
routes_ct9540	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step. Routes are polylines along ESRI's primary and secondary roads in McDowell County within each census tract. Seven routes were generated for this census tract.	routes_ct9540.zip
ct9540_new_addresses	HAI field workers	Shapefile / Feature Class	WGS 1984	New address points were separated from the assigned_stops_ct95xx feature class using the select by attributes function. New points were exported and reverse-geocoded to create XY coordinates and to verify street address data collected in the field. A total of 118 new addresses were collected in this census tract.	ct9540_new_addresses.zip
censustract_9542	United States Census Bureau	Shapefile / Feature Class	WGS 1984	Using partnership_shapefiles_21v2_54047, select by attribute each specific census tract polygon to create separate census tract	census_tract_9542.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				polygon for use in Route Planning analysis.	
censustract_9542_residences	McDowell County GIS	Shapefile / Feature Class	WGS 1984	Using mcdowell_R1-R6, select by location within census_tract_xxxx polygon, those residential points entirely within the census tract polygon to create separate point shapefile of residences within census tract. A total of 1,079 residential units in this census tract.	censustract_9542_residences.zip
assigned_stops_ct9542	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step above. Assigned stops are the residential point locations along each route within each census tract. A total of 1,114 residential units were verified or added in this census tract.	assigned_stops_ct9542.zip
routes_ct9542	HAI analyst	Shapefile / Feature Class	WGS 1984	Route planning was completed in ArcGIS Online using the residential points created from the select by location step. Routes are polylines along ESRI's primary and secondary roads in McDowell County within each census tract.	routes_ct9542.zip

Dataset name	Source organization	Format	Coordinate system / projection	Analysis steps	Data Access URL
				Six routes were generated for this census tract.	
ct9542_new_addresses	HAI field workers	Shapefile / Feature Class	WGS 1984	New address points were separated from the assigned_stops_ct95xx feature class using the select by attributes function. New points were exported and reverse-geocoded to create XY coordinates and to verify street address data collected in the field. A total of 21 new addresses were collected in this census tract.	ct9542_new_addresses.zip

2.2 Non-Spatial Data

Non-Spatial data were used or merged with geographic data for this project.

3 Software and Tools

3.1 GIS Software

Acknowledging that there are many different GIS software applications that can be utilized for geospatial analysis, ArcGIS Pro 3.6.3 was used for all viewing, editing, and analysis of geographic datasets received from key stakeholders. Shapefiles created from initial datasets received from stakeholders were transformed, packaged, and uploaded to ArcGIS Online for route planning, creation of field mapping surveys, and sharing purposes.

3.2 Supporting Software

As with most projects, Microsoft Office products were utilized for documentation and project planning activities. Address verification in the field was completed with ESRI's Field Maps mobile application. This mobile application was chosen over Survey123 because the Survey123 Field App is in the process of being deprecated in favor of a newer Survey123 version that has not been released yet.

4 Data Preparation

4.1 Data Quality Review

The original geographic data file received from McDowell County GIS was a layer file which included 12,986 commercial and housing unit point locations with Census Bureau structural classifications, street addresses, and brief descriptions of most structures. Data completeness checks, attribute validation and definitions were discussed in detail and completed with the McDowell County GIS Specialist during meetings with the SHAPE team, prior to any analysis.

4.2 Coordinate System Standardization

All layers received from McDowell County GIS were in GCS NAD 1983, as recommended by the West Virginia GIS Technical Center. No geographic transformations were applied to the original geographic layers. To prepare the data for route planning and surveying, the original structural unit layer file received from the county GIS Specialist was exported as a shapefile to access the spatial vector data. This shapefile served as the analysis file for route planning and surveying using ArcGIS Online Route Planner and Field Maps.

Field data collection specifically requires published and online hosted feature layers, requiring transformation to WGS 1984 Web Mercator Auxiliary Sphere. This transformation is applied automatically to the shapefile when it is imported into ArcGIS Online.

4.3 Data Cleaning

The original layer file received was structurally and geographically complete with only a few null values in the structure description attribute. The null values did not affect subsequent analysis, and no geometry had to be repaired or edited. Most attributes were retained for future analysis. To access the spatial vector data, the layer file received from the county was exported as a shapefile, and no data cleaning was necessary prior to, or after this step.

4.4 Data Integration

For this specific project, no data was joined, related or merged with any of the layers or shapefiles utilized for field data collection or subsequent analysis. However, for accurate route planning, location coordinates were necessary. Address data available in the original shapefile received from the county were geocoded with ESRI's World Geocoder Service and produced detailed location attributes including coordinates. The mean geocoding accuracy score for the entire county's 10,899 residential units was 99.57% including 10,846 residential points with a geocoding score of 95.48% or higher.

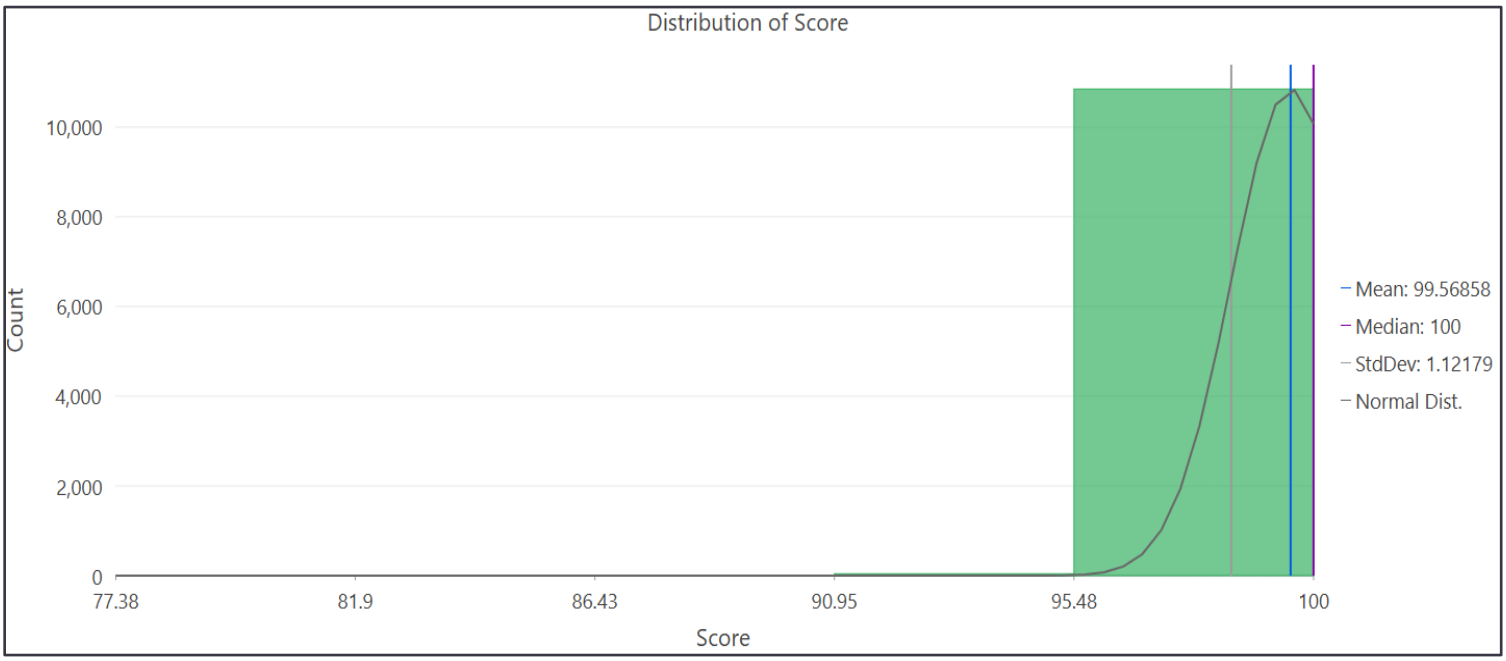
A very small number of geocoded points had scores of 90% or below. A summary of the geocoding results is in Table 2 and Figure 1 below.

Table 2: Geocoding results

Table 2: Geocoding results

Min of Geocoding Score	Max of Geocoding Score	Range	Count
77.38	81.9	77.38-81.9	3
81.9	86.43	81.9 - 86.43	5
86.43	90.95	86.43 - 90.95	1
90.95	95.48	90.95 - 95.48	44
95.48	100	95.48 - 100	10,846

Figure 1: Geocoding results



5 Spatial Analysis Methods

5.1 Study Area Definition

The study area was defined by the McDowell County boundary polygon sourced from the Census Bureau's TIGER / Line Shapefile data repository. All residential points received from the county and later assessed by the SHAPE team were included within this boundary.

The study area was further divided into 8 census tracts for route planning purposes, using the Census Bureau's Boundary and Annexation Files. For this pilot project, and upon recommendation from the Census Bureau, only five census tracts were surveyed. These census tracts are listed in Table 1 above.

5.2 Geoprocessing procedures, parameters and, methods

Analysis steps are described in chronological steps below. Other than ground truth verification described in section 6.1 below, no specific spatial analyses were performed.

5.2.1 Data cleaning

No specific data cleaning steps were performed on the original dataset with all point locations received from the McDowell County GIS Specialist. However, some data cleaning steps were performed in ArcGIS Online after field data verification to ensure addressing discrepancies were corrected. For example, most new address locations were added to the maps by field workers,

but no street address information was entered into the survey; some new street addresses entered in the datasets were incomplete, or some locations could not be surveyed due to geographic constraints.

5.2.2 Data delimitation

The original layer file received from the county consisted of 12,986 residential, public services, and commercial or business points. This dataset included an attribute that described each structural unit type as commonly implemented and utilized by state-and local municipalities' zoning boards. In West Virginia, it is not uncommon for counties or local municipalities to have differing definitions on residential zoning. The SHAPE team consulted with outside stakeholders on which structural types to include in the analysis data. After this step, it was decided that all points with residential designations R1 through R6 should be extracted from the dataset using the 'select by attributes' feature in ArcGIS Pro. Local definitions of the structural housing unit types are summarized in Table 3 below. This dataset consisted of 10,899 residential units and was exported as the residential points dataset and utilized as the analysis dataset of the project.

Table 3: Residential point dataset

Table 3: Residential point dataset

Residential type	Point type	Value	Sub. range	Notes	Addition information	Special considerations
Single Family	R1	790	80	A **single-family home** refers to a standalone residential structure designed to house one family or household. It is the most common type of residential property in many areas and is typically zoned as **residential.** **Single-family homes** often have commonalities. The structure is often a detached building with no shared walls that include private living areas such as bedrooms, bathrooms, kitchen, and common spaces. Occupancy wise, it is designed to house one household, though it may be owned or rented. These often include spaces like basements or garages that are part of the single household's use. **Single-family homes** are zoned as *residential single-family*, which often prohibits multi-family		**Subdivision Housing:** Single-family homes within planned developments or subdivisions may include shared amenities (e.g., pools or parks) but are still classified as single-family units. **Accessory Buildings:** Detached garages, sheds, or accessory dwelling units (ADUs) may accompany single-family homes but are considered secondary to the primary dwelling. (*If the Single-Family home is in fact an **Accessory Building**, consider using the point type **(A1) for Accessory Building**.*)

Residential type	Point type	Value	Sub. range	Notes	Addition information	Special considerations
				dwellings or commercial activities.		
Multi Family	R2	790	81	Multi-family housing refers to residential buildings designed to accommodate multiple separate households or families within a single structure or property. Each household typically has its own private living space, such as an apartment or unit, but some facilities and amenities may be shared.	*Duplexes, Triplexes, and Quadplexes: Small multi-family buildings with two, three, or four units, often resembling single-family homes. Units may be side-by-side, stacked, or a combination.* *Apartment Buildings: Larger structures with multiple units, ranging from small apartment buildings to high-rise complexes. Commonly found in urban areas.* *Townhomes: Multi-unit properties where units are connected side-by-side and share walls but often have separate entrances.* *Condominiums: Individual units but*	

Residential type	Point type	Value	Sub. range	Notes	Addition information	Special considerations
					share community facilities (e.g., dining rooms, recreational spaces).	
Mobile Home	R3	790	82	A **mobile home** is a prefabricated housing structure built on a permanent chassis, designed to be transported to a site and used as a residence. Mobile homes can serve as primary residences or temporary housing, and their classification depends on their use and location.	***Single-Wide:*** *Narrower and more compact.* ***Double-Wide:*** *Made of two joined sections, providing more space.* ***Park Models:*** *Smaller mobile homes, often used seasonally or recreationally. A* ***mobile home*** *on private land may be classified as a* ***single-family residence****. A* ***mobile home park*** *would fall under* ***multi-family residential*** *or* ***mixed-use****, depending on additional facilities.*	Things to consider when setting the classification on a mobile home is the residential usage and zoning. **Single-Family Residential Areas** are created when a Mobile Home is placed on privately owned land. If a Mobile Home is placed to be part of a mobile home park with multiple units, it is considered a **Multi-Family Residential**. (*Unless the mobile homes are assigned individual addresses.*)
Other Residential	R4	790	83	"**Other Residential**" is a catch-all category used to classify residential properties that do not fit traditional or common	***Specialized Residences****: Mobile homes or manufactured*	

Residential type	Point type	Value	Sub. range	Notes	Addition information	Special considerations
				classifications like single-family homes, multi-family units, or apartments. This category is typically used for unique or less common types of residential properties.	*housing (if not part of a mobile home park), Tiny Homes (on private lots or in communities), and Houseboats.* ***Mixed-Use Residential:*** *Buildings with both residential and commercial spaces (e.g., live-work units).* ***Shared Living or Co-Housing:*** *Group homes for unrelated individuals (e.g., assisted living facilities not classified as institutional).* ***Unique Residential Structures:*** *Converted barns, warehouses, or churches used as homes. Off-grid homes or Eco-Homes.*	
*Camp/Bungalow**	R5	790	84	A ***camp*** or ***bungalow*** typically refers to a small, modest structure used for recreational, seasonal, or temporary housing. These structures are	****A lakefront bungalow rented** to tourists may be categorized as **commercial lodging (CL)**. **A bungalow in a**	Emergency Services: Proper addressing is critical, especially for isolated camps or

Residential type	Point type	Value	Sub. range	Notes	Addition information	Special considerations
				often associated with leisure activities or vacation stays, and their classification depends on their purpose and zoning. For a Camp the purpose is often used for seasonal or temporary stays, such as summer camps, hunting camps, or fishing retreats. Camps can include clusters of small cabins or rustic shelters on a large property. The construction of a camp is often of simple structures, sometimes lacking modern utilities (*e.g., water, electricity*). Camps range from basic tent platforms to more developed cabins. As for a Bungalow, there purpose is to be a small, single story residential structure often used as a vacation or second home. Bungalows may also be a primary residence in some areas, especially in older or more compact neighborhoods. The construction of a Bungalow is typically a compact design, usually with a small footprint	*residential neighborhood would likely fall under residential zoning. A remote hunting camp used seasonally may be classified as recreational or temporary residential. The classification depends on the use case, location, and zoning regulations in the area.*	clusters of bungalows.

Residential type	Point type	Value	Sub. range	Notes	Addition information	Special considerations
				and often featuring porches or verandas. *Could be considered a commercial structure. Alternative information: POINT TYPE: CL , Sub. Range: 85		
Seasonal Home	R6	790	86	A **seasonal home** refers to a residential property that is occupied only during certain times of the year, often for recreational or leisure purposes. These homes are not primary residences but are used as secondary homes for vacations, holidays, or specific seasonal activities.	***Lake Houses:*** *Homes used during summer for recreation.* ***Ski Cabins:*** *Seasonal residences in winter sports destinations.* ***Snowbird Homes:*** *Properties in warmer climates used during winter months by individuals escaping colder regions.*	

5.2.3 Step 3: Data Cleaning

The residential unit layer with 10,899 points was further categorized into the county's eight census tracts using the select by location tool in ArcGIS Pro, with the county's census tracts serving as the polygon boundaries. This step divided and prepared the residential unit data points for route planning and subsequent analysis in ArcGIS Online. These datasets were uploaded as hosted feature layers to ArcGIS Online and automatically transformed from the geographic coordinate system NAD 1983 to WGS 1984 required by the software for accurate field data collection.

5.2.4 Step 4: Route Planning Analysis in ArcGIS Online

In addition to categorizing the residential point layers by census tract, route planning analysis required a start point. Starting points were created by selecting each public service site listed in the original layer file received from McDowell County GIS. Again, census tract polygon boundaries were used to select one public service site within each census tract, and the shapefiles were exported as hosted feature classes to ArcGIS Online. With input and confirmation from the Census Bureau, the SHAPE team planned address verification for five out of eight census tracts in McDowell County.

The route planning analysis tool in ArcGIS Online was used to plan address verification routes for the five census tracts, to ensure manageable driving routes for field workers. Due to some software limitations, the number of routes within each census tract were balanced with the number of residential unit points, where a maximum of 200 points per driving route (excluding the start and end points) were allowed. Each field verification route started at a single public service site in each census tract, following ESRI's primary and secondary roads. The planned routes produced were visually verified against a road layer received from McDowell County GIS, to ensure that the planned driving routes were as close as possible to locally verified routes.

The route planning tool produced sequentially numbered assigned stops from the residential unit layer in each census tract, with polyline routes following primary and secondary roadways in the county. It should be noted that the naming of the assigned stops and routes is a default setting in the route planning software. Examples of routes with assigned stops and census tract boundaries are in Figures 2 and 3 below

Figure 2: Example of routes with assigned stops and census tract boundaries

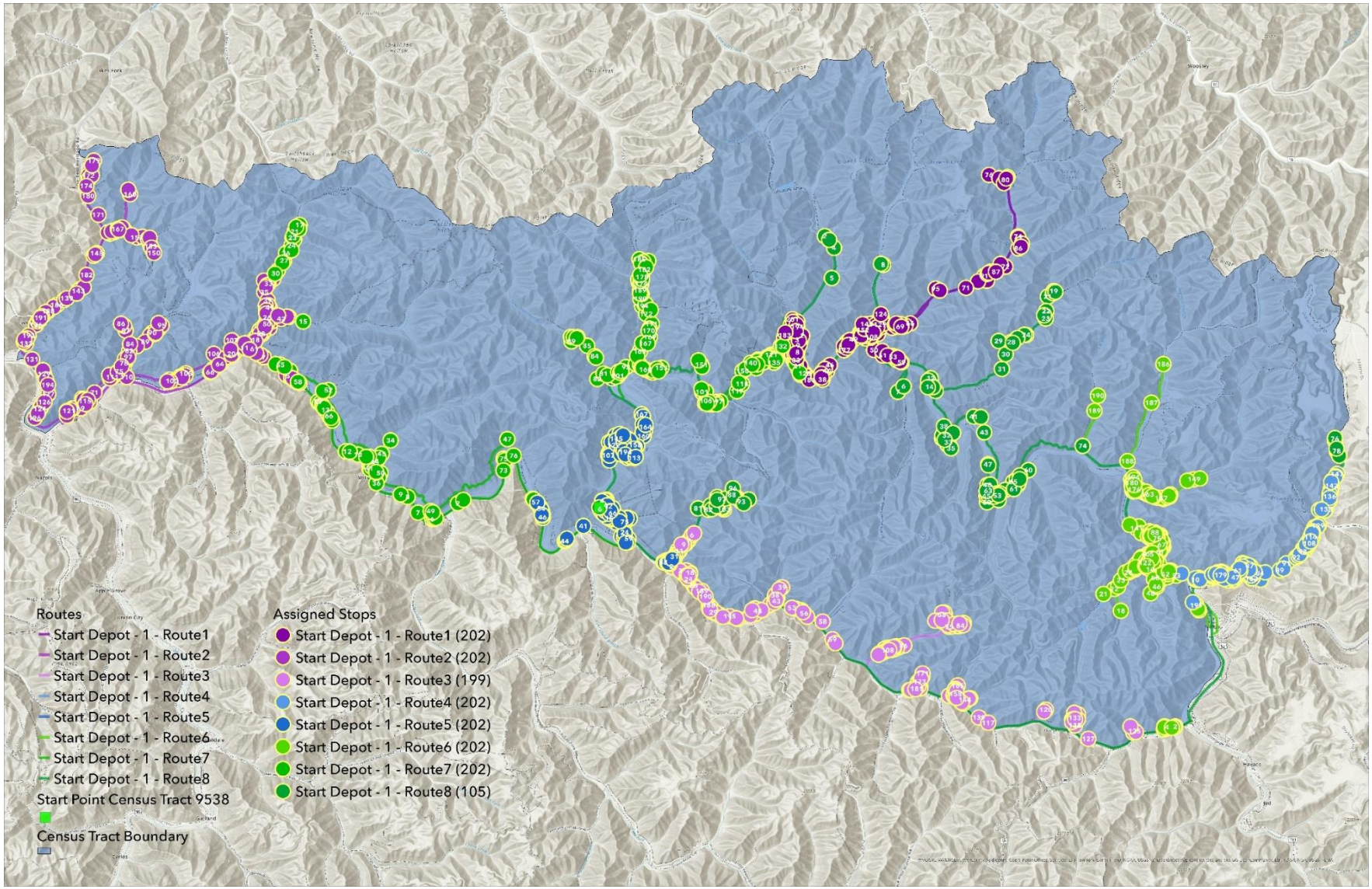
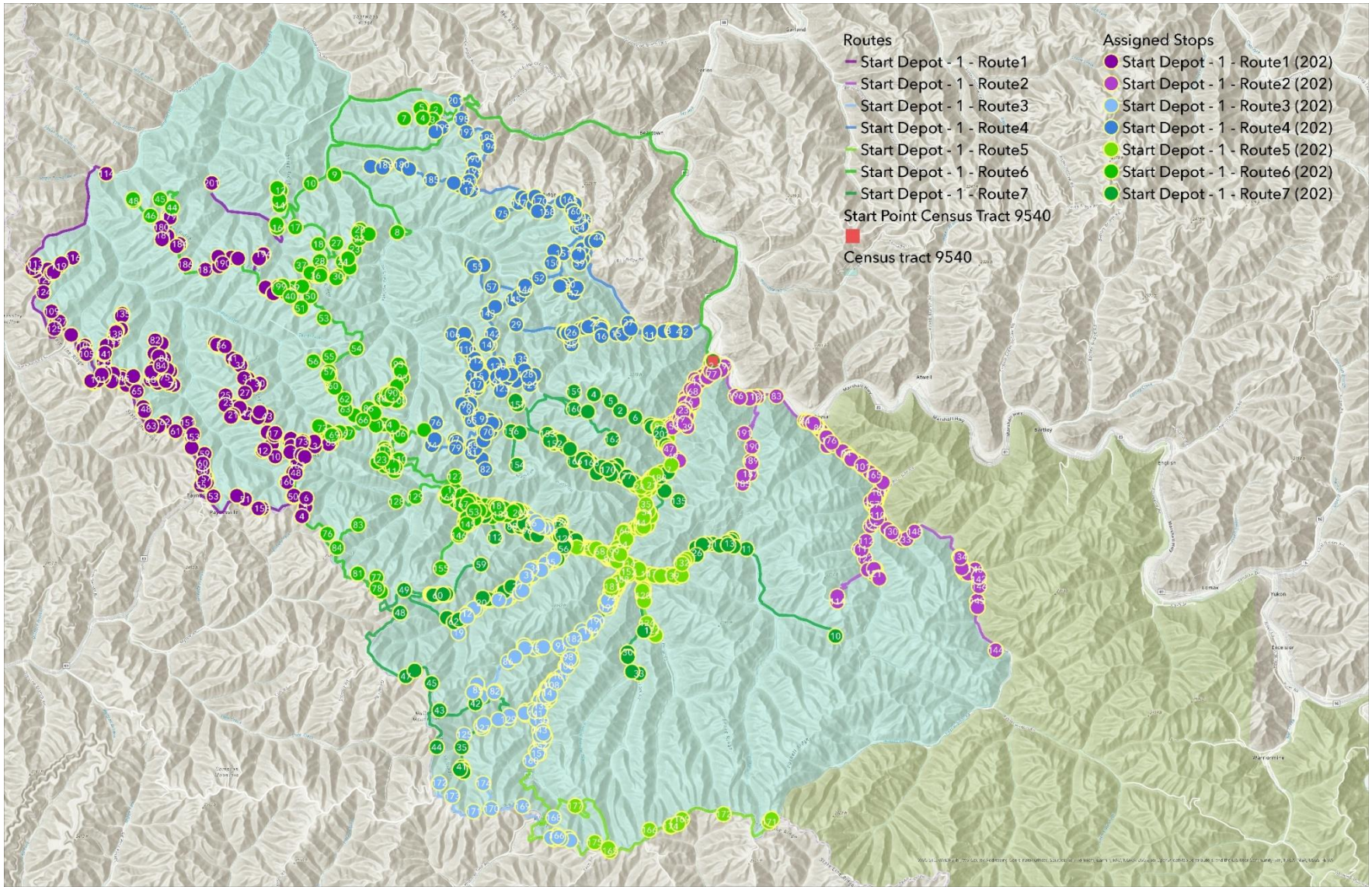


Figure 3: Example of routes with assigned stops and census tract boundaries



5.2.5 Step 5: Data set-up and configuration

The assigned stops point layer for each census tract was used to survey the addresses in the county. Web maps consisting of the assigned stops (residential units layer) and census tract boundaries for each of the five census tracts were created and saved as public data collection feature layers. Additional dataset configurations in each web map included:

- Full geometry and text editing access for users
- Data deletion privileges were removed to prevent unintentional removal of original data points and web maps
- Change tracking enabled
- Online and offline syncing capability enabled
- Searchable in ArcGIS Online with the phrase 'Surveying Census Tract 95xx'

Example of web maps provided in Figure 4 and 5 below:

Figure 4: Example of web map

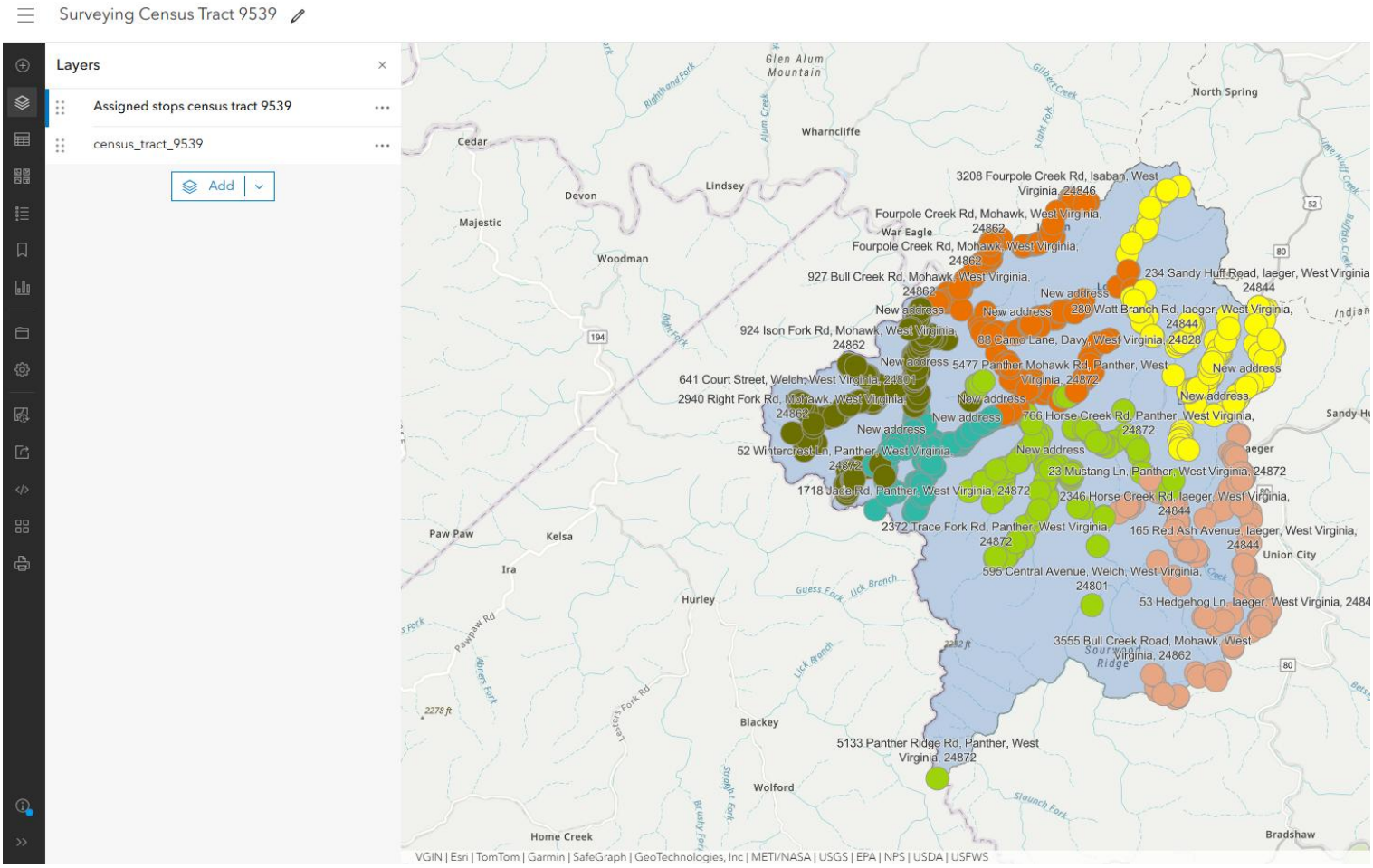
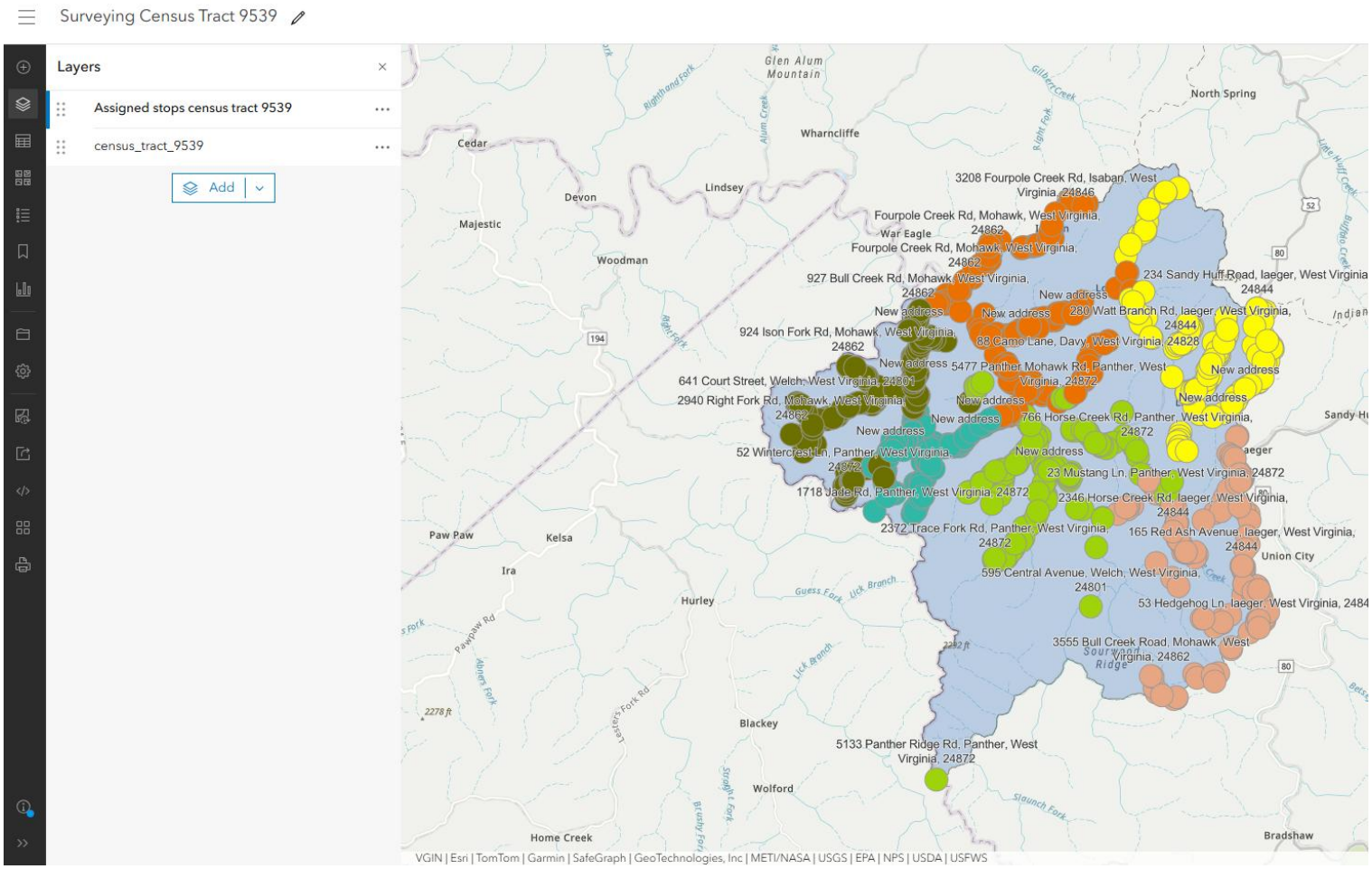


Figure 5: Example of web map



5.2.6 Step 6: Building and configuring the field survey in ArcGIS Online

For this specific pilot project, fully hosted and correctly configured web maps consisting of publicly available hosted feature layers were required to create field surveys in ESRI's Field Maps Designer. Although field surveying tasks can be assigned to specific ESRI users, the SHAPE team designed the survey without this functionality due to time constraints.

Using the assigned stops point layer in each census tract, the field survey form was created in ESRI's Field Maps Designer application. This application connects with a mobile field data collection application with numerous geographic capabilities for fast, accurate field data collection. As with the commonly used Survey123 application, the Field Maps Designer application is highly configurable, with commonly used field data collection capabilities, including photo and document attachments, drop-down menus, selection buttons, text boxes and data contingencies.

The following configuration settings were applied to each census tract in the Field Maps Designer application:

- Offline data collection turned on for each layer, in both the web map and in the field survey application.
- Ensure a basemap that can be used offline is selected. For this specific project, ESRI's World Topographic Map loaded in the mobile application without issue.
- Creation and updating of the offline map area.
- All features and attachments in the field survey were selected to be synced both offline and online on the mobile device. This setting can be customized based on need.
- Tables and standalone tables were not made visible in the mobile application. This setting can be configured to show all tables in the survey.
- Due to geographic constraints, service area, and mobile device limitations, the location accuracy was set at 50 feet, after consultation with ESRI technical support representatives. The default setting in the Field Maps Designer application is 32 feet and is likely more realistic in areas without mobile service access issues.
- GPS averaging and streaming settings were turned off.
- Snapping was allowed.
- New features were set to be created from the location of an existing feature, place, or dropped pin. All attributes were set to be copied when a new point was added. This ensured that field workers could document new residences along routes.
- Layer filters in the field data collection application were turned off.
- Depending on specific organizational regulations, location tracking of each field worker logged into the application can be turned on.
- The online map extent was set to default, with the manually created map area available for offline field data collection.
- Search was enabled by address, in the mobile Field Maps application.
- The sharing level in Field Maps Designer was set to 'Public'.

- Along with public sharing, each field survey can be accessed by a unique link or QR code.

Although the above settings seem complex and time-consuming, the Field Maps Designer application is user friendly and intuitive.

5.2.7 Step 7: Testing and reconfiguring the field survey

Due to inexperience with designing field surveys using the Field Maps Designer, some technical difficulties were encountered. However, if organizational account access includes technical support, it is important to establish a support connection early in the project to ensure efficient set-up and access to troubleshooting in the field.

After each census tract survey was set up in the Field Maps Designer, SHAPE team members were assigned ESRI field worker accounts and tested the mobile Field Maps application by adding dummy locations and entering fake data into the survey. Based on their feedback, the surveys and maps were adjusted or edited to allow easy download of offline map areas, and easy access to each census tract survey on mobile devices. It is important to allow enough time for this stage, prior to field data collection operations.

6 Quality assurance and validation

6.1 Accuracy assessment

Ground truth verification was initiated by layering field collected addresses over the addresses in the original layer file received from McDowell County GIS. This was completed in R using the sf package, which contains functions for manipulating and analyzing geographic data. Numerical validation was also complete. Data was merged by address and Euclidian distance between the two GPS locations was calculated. Newly identified addresses do not have an initial comparison point and addresses marked as no longer livable or vacant were also removed. Data was also separately merged by a slightly truncated GPS location to inspect discrepancies between addresses. Truncation was necessary due to field collection limitations.

Using field collected variables and user comments, vacant addresses and unlivable addresses were flagged. Newly identified addresses were also flagged. Data was combined and binned by block group using the st_join function with the appropriate shapefile. From the stringr package, the str_detect function was used in R to identify the word “vacant” in user comments. Addresses marked as unlivable during data collection were filtered out of the binned 2026 field collected data.

6.2 Comparison with reference datasets

Multiple data sets were used as references for comparison. The two most notable were the American Community Survey (ACS) five-year estimated occupied housing counts by block group (B25002) and the 10-Year Census counts by census tract. ACS data was used for more precise geographic comparisons while census counts were used for more broad comparisons.

Counts of non-vacant and livable field collected addresses by block group were compared to ACS occupied address counts. These comparisons provided insight into relationships between geography and address density accuracy. Counts of non-vacant and livable field-collected addresses by census tract were compared to Census Bureau occupied address counts. This provided insights into more broad relationships geographically while still being more informative than county level estimates.

6.3 Error rates

Point to point comparisons from the field collected address locations and the original layer file received from McDowell County GIS allowed RMSE (root mean squared error) of the Euclidean distance to be calculated. This value was ~ 4e-4 m, which is potentially so low partially due to field collection limitations and the initial precision of the original layer file.

6.4 Peer review / QA procedures

6.4.1 Internal review process:

- Housing unit data received from the county were reviewed with the GIS specialist prior to any analysis. Review actions included:
    - clarification of attributes with the GIS specialist at the county
    - requesting a data dictionary
    - geocoding of address data to ensure location and address data accuracy
- Geocoding of the address data received from the county also created XY coordinates and a geodatabase file needed for route planning.
- After all addresses in five out of eight of McDowell County's census tracts were verified during the field work phase, the team cleaned the data prior to sharing with stakeholders. Data cleaning actions for this phase of the data review process included:
    - Verifying all attributes in the field survey were as complete as possible

- Ensuring that any new address locations had the correct street address in the survey attribute titled 'If you answered 'No' in the previous question, or of this is a new property, enter the correct street address above'
- General location accuracy for new addresses such as placement of the point on the correct structure
- Ensuring each census tract shapefile was in the correct projection for online field data collection (WGS 1984)

6.4.2 Independent verification:

The SHAPE data underwent an independent verification process through our partners at the Census Bureau, providing an additional layer of quality assurance for the project's findings.

Our validated field maps, including newly identified addresses and housing locations, were compared against the Census Bureau's existing address information and available aerial imagery. This independent comparison helped confirm the accuracy of SHAPE's field-validated data, identify areas where newly identified locations were not reflected in existing address lists, and strengthen confidence in the methodology and findings.

6.5 Limitations

The original dataset received from McDowell County was very detailed and complete. However, it was shared as a read-only layer file without access to the geographic metadata, as with a shapefile or geodatabase. Due to this limitation, the SHAPE team geocoded the residential addresses against ESRI's World Geocoder to create geographic metadata necessary for route planning and surveying. Geographic error testing thus occurred between geocoded data and field data collected by the SHAPE team.

Due to inexperience with the set-up and configuration of the field data collection application, data collection settings created duplicate address points for surveyed or verified addresses. Set-up and configuration had to be balanced with protecting geographic integrity of original data received from the county and allowing geometric and attribute creation for any new residential units encountered in the field. This created some duplicates for anomalous residential units in the field data, based on field workers' observations.

Other data limitations include unverified addresses due to access or roadway issues and some subjective uncertainty about residential unit occupation due to leaf cover.

7 Mapping and visualizations

7.1 Cartographic standards

Cartographic standards are the way geographic data are shown on a map. Simple symbology with bright color schemes and basic shapes (circles) were configured during route planning analysis. Residential units were classified by routes, using grouped color schemes matching the routing color scheme as seen in Figures 1 and 2. Each residential point was labelled with the routing sequence number, i.e., Each starting point on the route is 0, and the assigned stops labelled with the routing sequence along the route.

For field data collection purposes, different, brighter color schemes were applied to residential units on planned routes in each census tract and labelled with the geocoded street address in the Field Maps Designer application as seen in Figures 4 and 5. These color schemes allowed for visual differentiation between census tracts in the mobile field data collection application.

7.2 Map production

To ensure accurate field data collection, points and boundaries were configured to be visible at all mapping scales, with building footprints being the smallest scale. Map legends included routes, polygon boundaries (census tracts), and residential units grouped by routing classification, i.e., residential units along Route three in Census Tract 9539 is orange, and residential units along Route five in census tract 9540 is light blue in the Field Maps application.

8 Outputs and Deliverables

8.1 GIS Data Products

All feature classes used for field data collection and analysis created by the SHAPE team are listed in Table 1 of this document. Geodatabases are included in the feature classes and serve as the geographic metadata repositories for all layers. No raster features were utilized in this project.

8.2 Maps

Web maps were created at each stage of this project and used for different purposes such as sharing, storing original data, and field data collection purposes. Route planning analysis for each census tract was completed prior to uploading data to the Field Maps Designer application. Web maps created for this project are listed in chronological order beginning with the SHAPE database and ending with Surveying Census Tract 9542 in Table 4 below:

Table 4: Web maps created for the SHAPE project

Table 4: Web maps created for the SHAPE project

Web map title	Purpose	Visible feature classes	Links
SHAPE Database	Share data with stakeholders and storing original data received from the county	ABS addresses, all residential units in the county, McDowell County road network, census tracts from BAS files	https://arcg.is/1GOXiy5 and https://arcg.is/1KOHCT2
Route Plan Census Tracts 9536, 9538, 9539, 9540, 9542	Route planning analysis and creation of separate residential unit and route layers for subsequent analysis	Public service site (start point), assigned stops, routes, census tract boundary	https://arcg.is/WfGHq https://arcg.is/0Km8DD1 https://arcg.is/vHXy80 https://arcg.is/1fm5Ov2 https://arcg.is/151aL04
Surveying Census Tracts 9536, 9538, 9539, 9540, 9542	Assigned stops created in route planning were uploaded into respective Field Maps Designer applications for surveying. Census tract boundaries were added to assist field workers	Assigned stops (residential units along routes in each census tract) and census tract boundary	https://arcg.is/0Xnu9r0 https://arcg.is/SyDq91 https://arcg.is/bDTOD1 https://arcg.is/ne0WW1 https://arcg.is/0bvnqv2

9 Results

9.1 Summary Statistics

Comparisons between census occupied address counts, ACS occupied estimations, original GIS layer address counts, and field collected occupied address counts by census tract provided insight into potential improvements. The percent improvement between the census counts and the field collected counts provides the most useful comparison since ACS tends to underestimate rural block groups. The primary limitation of the census counts is their availability by census tract instead of block group, which would provide more nuanced geographic insight. Census tracts in bold script were surveyed by the SHAPE team.

Table 5: Summary Statistics

Table 5: Summary Statistics

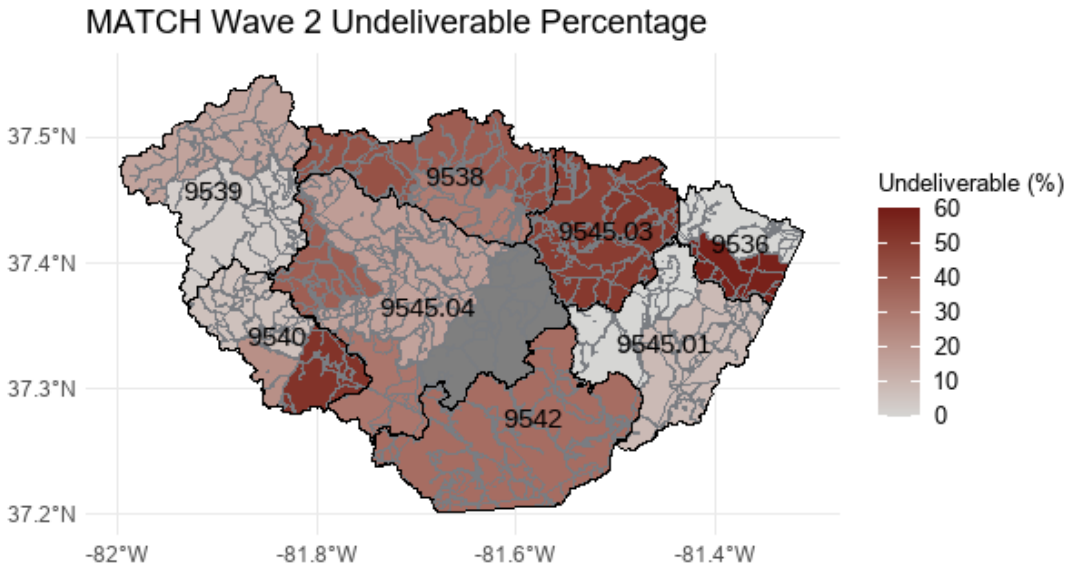
Census Tract	Census Occupied Households	ACS Estimated Occupied Households	Original GIS Household Counts	Field Collected Occupied Households	Percent Improve (%)	RMSE of Euclidean Distance (m)
954503	1,337	1,163	1,570	NA	NA	NA
954501	725	699	990	NA	NA	NA
954504	1,624	1,380	2,351	NA	NA	NA
9540	791	665	1,343	1,254	58.5	0.00041
9542	696	661	1,042	1,029	47.8	0.00041
9539	874	790	1,038	1,106	26.5	0.00041
9538	1,050	750	1,436	1,438	37.0	0.00041
9536	386	359	611	708	83.4	0.00042

The most improved census tract is 9536 which has shown an increase in occupied address counts by 83.4% when comparing the field counts to the census counts. Census tracts 954501, 954503, and 954504 were not surveyed based on recommendations from the Census Bureau. However, the original GIS layer file indicates potential undercounting in those tracts as well.

9.2 Analytical findings

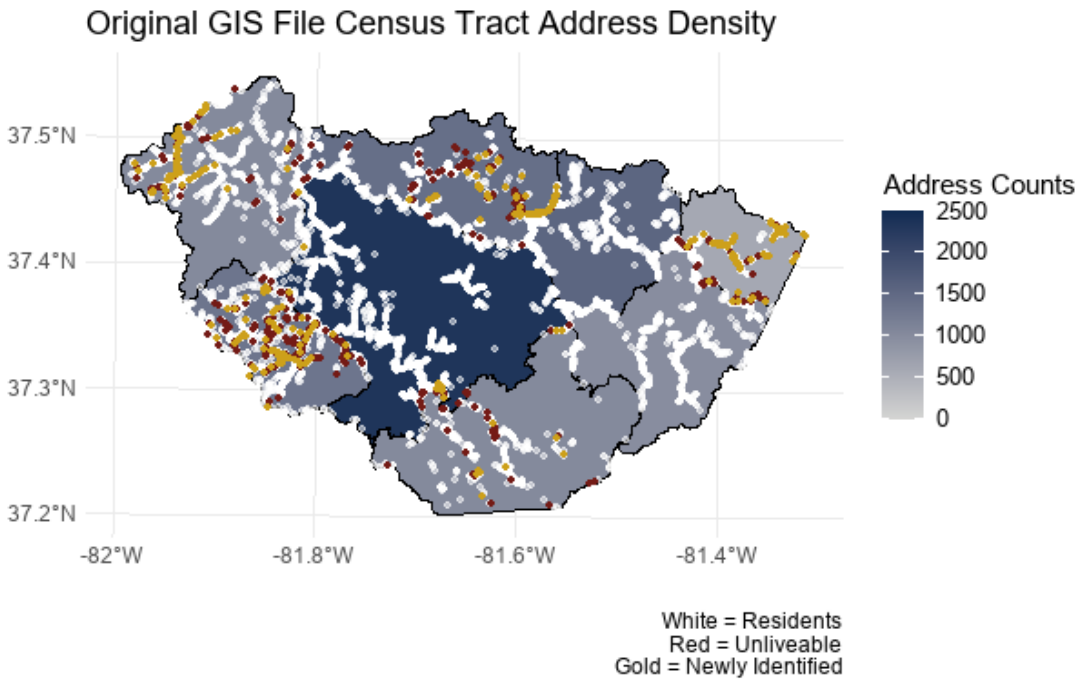
Several interesting trends can be observed in the mapping of various data sources.

Figure 6: Mountain State Assessment of Trends in Community Health (MATCH) undeliverable map



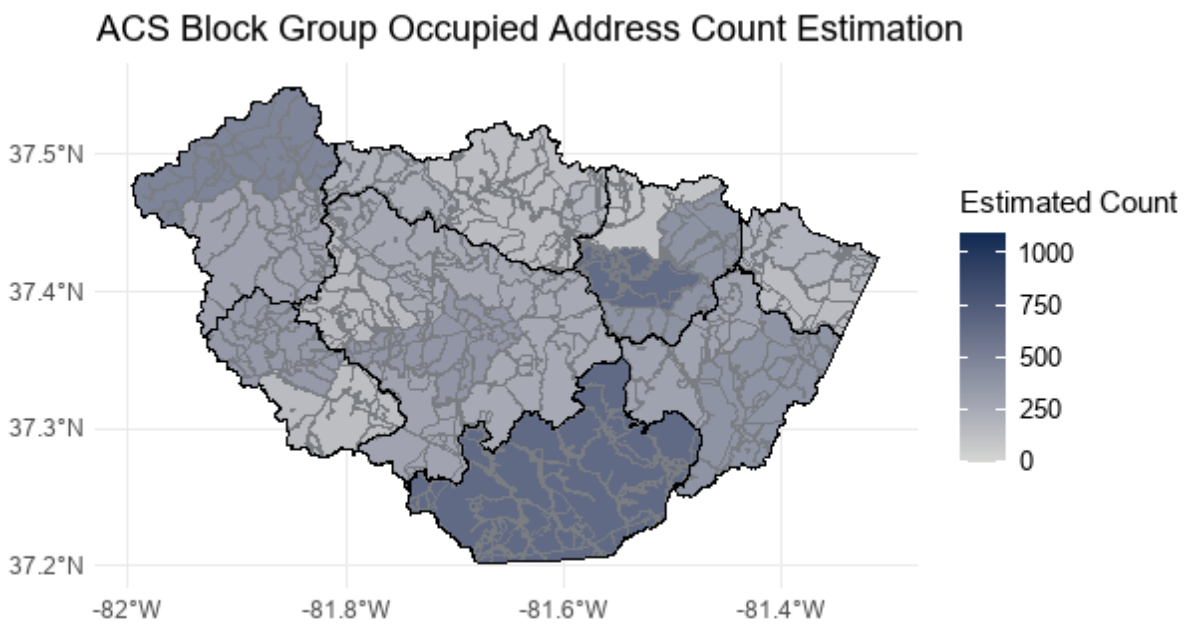
McDowell County had the highest undeliverable rate during wave 2 (2021-2022) of the HAI MATCH survey. This observation initiated the development of the SHAPE program.

Figure 7: Original GIS file address density map



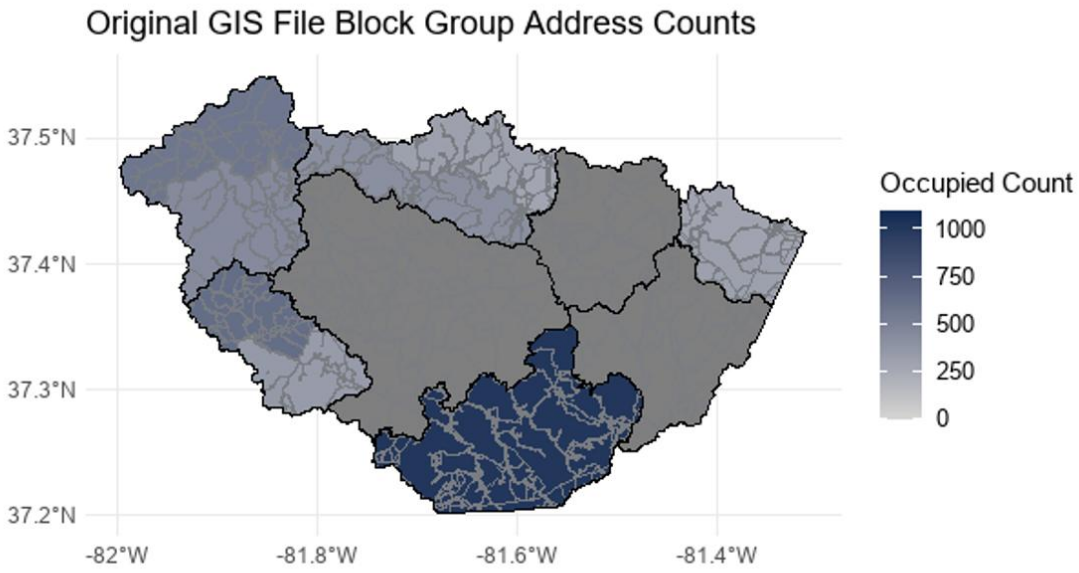
The SHAPE project focused on five specific census tracts with direction from the Census Bureau. Newly identified addresses in these tracts can be seen as individual gold dots. Addresses newly identified as unlivable or vacant are included in red. White dots represent initial addresses provided to the SHAPE team by GIS experts in McDowell County via the original GIS layer file. Locations where more gold than red dots can be observed potentially indicate tracts that are growing or have new development when compared to tracts with more red dots than gold.

Figure 8: ACS occupied address estimation map



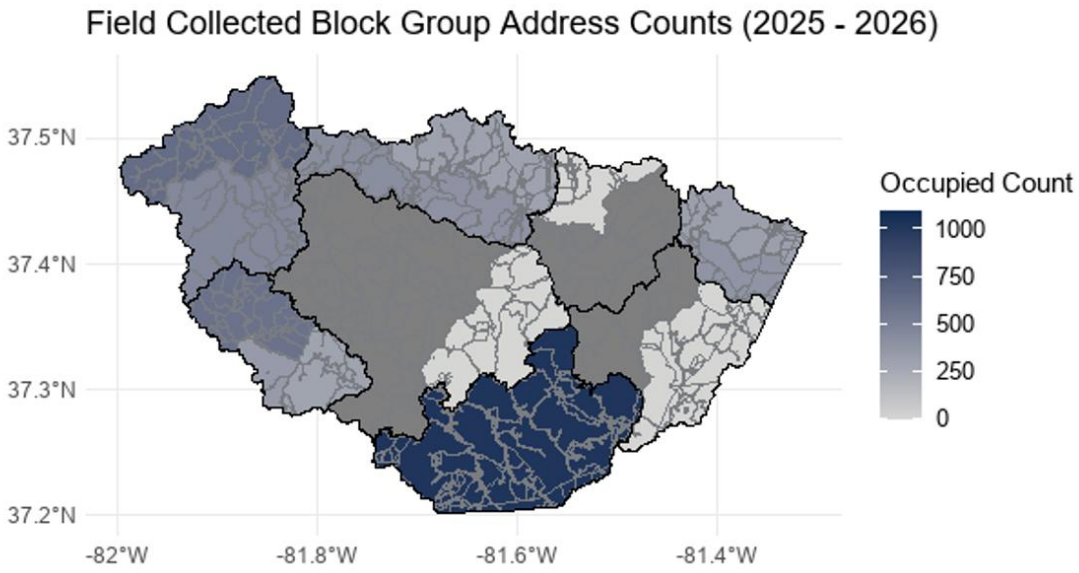
The ACS estimates for occupied households provides insight into population distribution in McDowell County. Note the scale for counts in this figure and Figure 9 and Figure 10 have the same legend scale.

Figure 9: Original GIS address counts map



Field-collected counts provide more accurate estimates when compared to ACS. Five block groups in McDowell County can be seen above with data provided by county GIS experts. The higher counts here potentially provide evidence of ACS underestimation, but field counts would be ideal for validation of this trend.

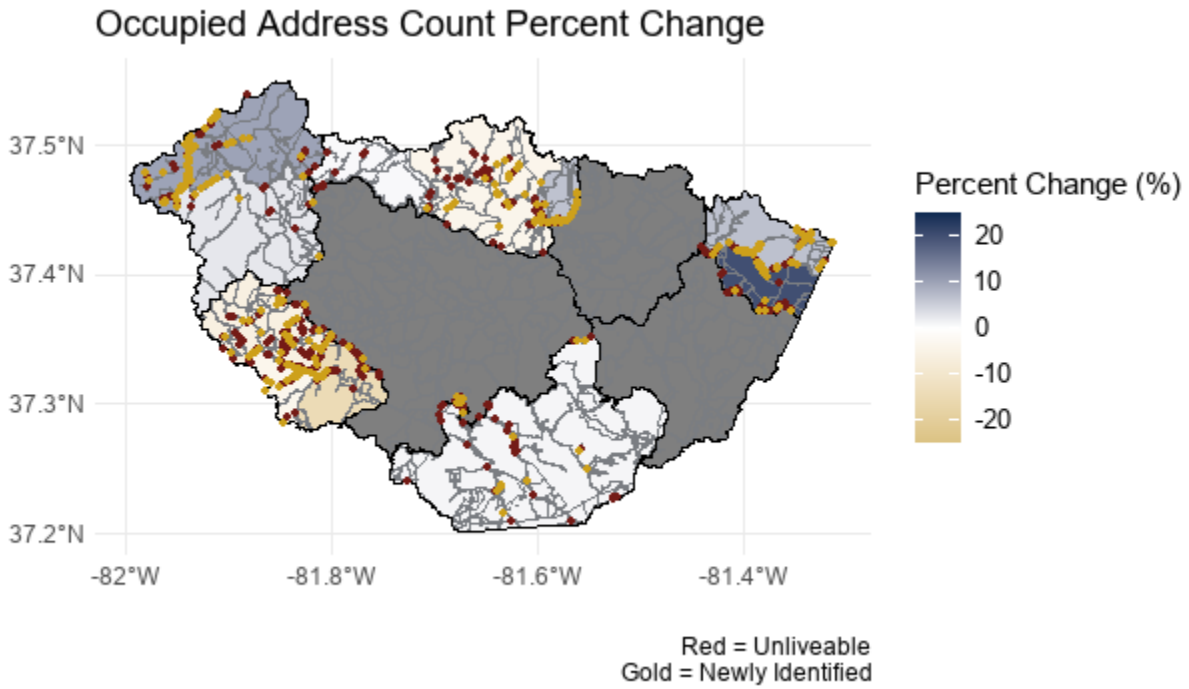
Figure 10: Field collected address counts map



HAI SHAPE team conducted field observations during 2025 and 2026 to provide a more updated occupied address count. When compared to ACS estimations and the original GIS file counts, evidence of ACS underestimation can likely be confirmed.

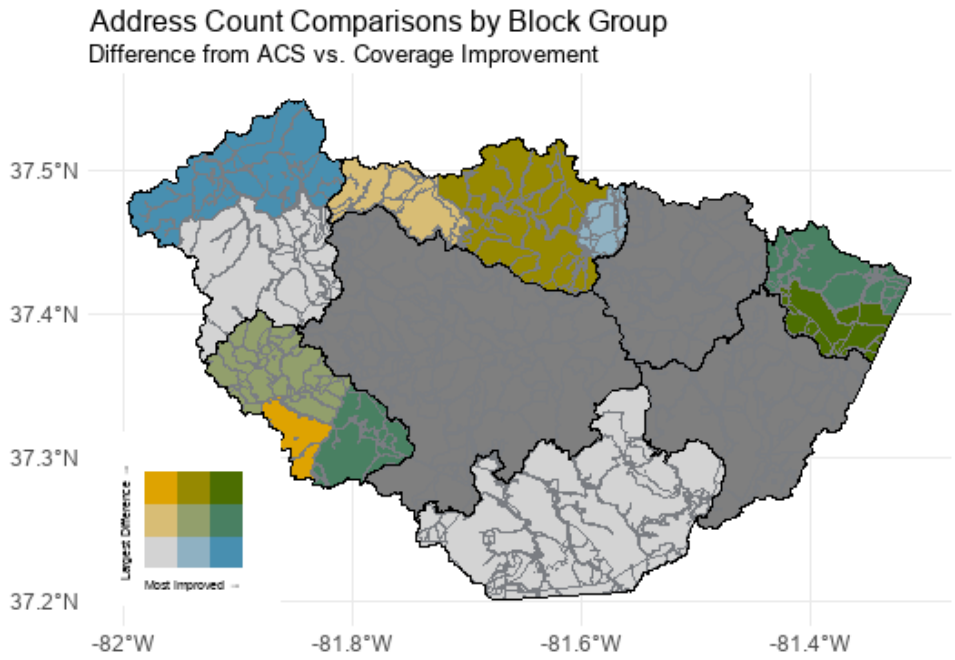
There are three block groups visible in Figure 10 in light grey. These are likely newly identified addresses that field observers noted to be in one block group during data collection but were in a different block group once data analysis was performed. They are included here to express a limitation of field observations since population density in these rural locations often lies on census boundaries.

Figure 11: Address counts percent change map



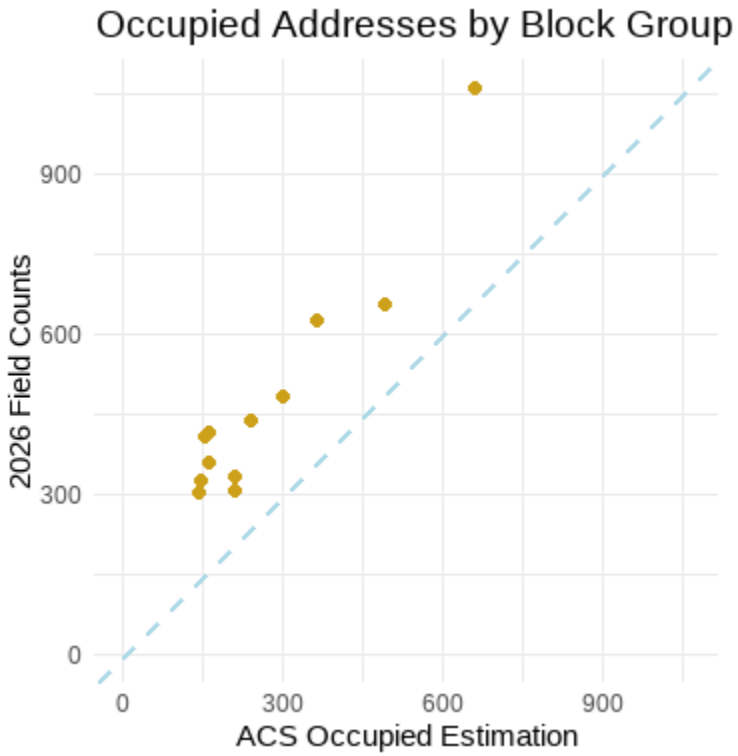
The percent change between the field collected address counts and the original GIS file shows an increase in counts for several measured block groups, though a few block groups do show a decrease. This decrease could be indicative of population movement or housing maintenance negligence. Interestingly, the block group with the largest increase and the block group with the largest decrease were the two block groups with the highest undeliverable rate in the MATCH wave 2 survey.

Figure 12: Bivariate choropleth comparison map



The improvement in coverage was calculated as the absolute value of the percent change in Figure 11. This allows for a more straightforward interpretation of coverage improvement since the absolute value of the change in address counts is indicative of a more accurate final count. Even if the final count is lower than the initially provided counts, the final lower count is a more accurate representation of the current counts for that block group and therefore shows an improvement in coverage. Comparing the improvement in coverage of the field counts to the estimates from ACS can identify a few different trends. Gold block groups indicate highest underestimation by ACS while having the least improvement in accuracy due to recent field counts. Dark green block groups indicate a combination of high ACS underestimation and improved field measurement accuracy. Blue block groups indicate improved field measurement accuracy with the lowest amount of underestimation, even though underestimation is still present. Light grey block groups indicate a closer ACS estimate with the least improvement from recent field counts, though it could also be indicative of a need for more investigation.

Figure 13: ACS estimation compared to field counts



The scatter plot in Figure 13 provides useful insight into the ACS underestimations of all block groups in McDowell County, confirming trends identified in the previous figures.

10 References

The following geographic sources were used during the creation of all maps in this project:

United States Census Bureau. (2025, December 23). *Tiger/line shapefiles*. Census.gov. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

Census Bureau Geographic Partnerships (2026, January 14). *Shapefile batch download - West Virginia*. United States Census Bureau. https://www.census.gov/geo/partnerships/pvs/partnership25v2/st54_wv.html

McDowell County GIS Department. *McDowell_Sites.lyrx* (2025, October).

U.S. Census Bureau. (2024). *Table B25002: Occupancy Status [2024 ACS 5-Year Estimates Detailed Tables].* Retrieved from U.S. Census Bureau Data Platform.

West Virginia GIS Technical Center. (2024). *Census Tracts with Population Data (2020) [Vector GIS Dataset, filename: CensusTracts_Census_2020_UTM83]. Derived from U.S. Census Bureau TIGER/Line Shapefiles and Decennial Census Data. Morgantown, WV: West Virginia University.* Available from the West Virginia GIS Data Clearinghouse: https://wvgis.wvu.edu/data/dataset.php?ID=427

Mueller, J. Tom and Santos-Lozada, Alexis R. 2022. *"Rural Population Research is Heavily Constrained by Federal Data Availability and Reliability."* Brief # 20228. Rural Population Research Network.

The following software, software components, and content were used for geocoding, route planning and surveying:

Esri. (2025). *ArcGIS Online* [Web Maps]. Environmental Systems Research Institute (Esri), Redlands, California, USA.

Esri ArcGIS Pro 3.6.3. (2025). *ArcGIS StreetMap Premium North America* [Geocoding]. Environmental Systems Research Institute (Esri), Redlands, California, USA.

Esri. (2025). *ArcGIS Online* [Route Planner]. Environmental Systems Research Institute (Esri), Redlands, California, USA.

Esri. (2025). *ArcGIS Online* [Field Maps Designer]. Environmental Systems Research Institute (Esri), Redlands, California, USA.

Esri. (2025). *ArcGIS Mobile* [Field Maps]. Environmental Systems Research Institute (Esri), Redlands, California, USA.

Pebesma, E., 2018. *Simple Features for R: Standardized Support for Spatial Vector Data.* The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With Applications in R (1st ed.)*. Chapman and Hall/CRC. Spatial Data Science | With Applications in R | Edzer Pebesma, Roger B

Prener C, Grossenbacher T, Zehr A (2022). *_biscale: Tools and Palettes for Bivariate Thematic Mapping_. R package version 1.0.0*, <https://CRAN.R-project.org/package=biscale>.

Walker K, Herman M (2024). *_tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames_. R package version 1.6.2,* <https://CRAN.R-project.org/package=tidycensus>.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). *"Welcome to the tidyverse." Journal of Open Source Software_,* *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

Wickham H, Miller E, Smith D (2022). *_haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files_. R package version 2.5.1,* <https://CRAN.R-project.org/package=haven>.

Wilke C (2024). *_cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'_. R package version 1.1.3,* <https://CRAN.R-project.org/package=cowplot>.

Appendix A: GIS Workflow Diagram

