

A Longitudinal Analysis of Property Values in Lubbock, 1945–2021: The Persistent Impact of Redlining

Contents

Section	Title
	Project Objective
	Introduction
	Chronological Development
1	Data Origin and Collection (Archival Years: 1945, 1975, 1985)
1.1	Archival Datasets — Description & Exploratory Analysis (1945, 1975, 1985)
1.1.1	Dataset characterization
1.1.2	Valuation fields used in EDA
1.1.3	Regional coverage indicators
1.1.4	Figures and interpretations
1.2	Contemporary Datasets — Provenance & Exploratory Analysis (2012, 2020, 2021)
1.2.1	File structure and address diagnostics
1.2.2	Valuation components and internal consistency
1.2.3	Distributional summaries (nominal USD)
1.2.4	Structure age (YearBuilt)
1.2.5	Neighborhood coverage (NbhdCode)
1.2.6	Figures and interpretations (contemporary years)
2	Methods: Data Cleaning, Address Standardization & Merging
2	Overview of the pipeline
2.1	Iteration 1 — Regex & Token Standardization
2.1.1	1945 workflow (irregular strings)
2.1.2	1975 workflow (light normalization + locality tokens)
2.1.3	1985 workflow (ZIP→city enrichment)
2.1.4	Case policy
2.1.5	1945 Pattern Library
2.2	Iteration 2 — GeoPy-based normalization (archival years)
2.2.1	1945 — preprocessing, geocoding, and retry logic
2.2.2	1975 — preprocessing, locality hints, and validation
2.2.3	1985 — ZIP-anchored hints and staged retries
2.2.4	Post-geocoding normalization(when present)
2.2.5	GeoPy Iteration outputs
2.2.6	Summary tables
2.3	Iteration 2.1 — Additional path: libpostal parsing + Nominatim standardization
2.4	Iteration 3 — Google Geocoding API canonicalization (archival years)
2.4.1	Acceptance & error handling
2.5	Cross-Year Diagnostic — Matching
2.6	Handling repeated addresses within a year (multi-unit buildings)
2.7	Archival merge — 1945, 1975, 1985 (post-Iteration 3)
2.8	Contemporary years — standardization & geocoding (2012, 2020, 2021)

2.9	Cross-Era Matched Merge (1945/1975/1985 ↔ 2012/2020/2021) — Lean Checkpoint
2.10	Six-Year Union Panel — Consolidation & Harmonization
2.10.1	Inputs and key set
2.10.2	Sequential left joins (string equality)
2.10.3	ZIP extraction (two-pass refinement)
2.10.4	Coordinate unification & cleanup
2.10.5	Output structure and exports
2.10.6	Presence reading and a simple trajectory
3	Spatial Augmentation in ArcGIS Pro — Layers, Assumptions, and Prepared Classifications
3.1	External expertise and working assumption
3.2	Layers imported/prepared in ArcGIS Pro
3.3	Georeferencing and alignment
3.4	Event layer and extents
3.5	Proximity structure around the possible redlining zone
3.6	Council districts
3.7	Prepared classifications added to LBK6Y points
4	Analytical Findings (with Zoning Context)
4.1	Study Frame and Inventory
4.2	Data Completeness and Readiness
4.3	Spatial Coverage — Districts × Zones
4.4	Value Structure and Cross-Year Correlations (diagnostic)
4.5	Property Value Levels by District (outlier-trimmed)
4.6	Property Value Levels by Redlining Proximity (outlier-trimmed)
4.7	Growth Dynamics by District (percentage change in adjusted means)
4.8	Parcel Price per Square Foot (PPFS)
4.9	Quality Notes and Analytical Guardrails
4.10	Executive Synthesis
5	Per-Square-Foot (PPFS) Metrics — Construction, Imputation, and Findings
5.1	Data inputs and PPFS definition
5.2	Source caveat — archival “area” is not a usable measurement
5.3	Indicator rows for year-membership
5.4	Lot-size imputation policy (archival years only)
5.5	PPFS availability and inflation to constant dollars
5.6	Distributional shape and citywide drift (2012–2021)
5.7	Spatial contrasts — PPFS by redlining proximity (2021)
5.8	District cross-sections (brief)
5.9	Quality guardrails and edge cases
5.10	Takeaways
6	Main PPFS Analysis (Nominal and Inflation-Adjusted to 2025\$)
6.1	Cohorts and Coverage
6.2	Definition and Units
6.3	Citywide Levels (2025\$) — All vs Complete

6.4	Nominal \$ Benchmarks (single-year interpretation)
6.5	Distribution Shape and Tail Diagnostics
6.6	Zone Gradient — 2021 (2025\$)
6.7	PPFS Growth
6.8	Additional Visual Diagnostics
6.8.1	What these figures show
6.8.2	Conclusions
7	Persistent Effects of Jim Crow-Era Redlining on Lubbock Property Values
7.1	Era-Appropriate City Limits: Source, Rule, and Integration
7.2	Possible Redlining Zones (RL1, RL2, RL3)
7.2.1	RL1 (1923 ordinance geometry)
7.2.2	RL2 (highway-era geometry)
7.2.3	RL3 (expert consultation)
7.2.4	Buffers for RL1–RL3
7.3	Analytical Framework and Cohorts
7.3.1	Inputs and provenance
7.3.2	Key variables and naming conventions
7.3.3	Cohort construction
7.3.4	Transformations and exclusions
7.3.5	Statistical testing and notation
7.3.6	Box Plots of three cohorts
7.3.6.1	RLs vs Cohorts
7.3.6.2	Groups (in vs out) vs RLs
7.3.6.3	Summary
7.4	Parcel Coverage and Counts by Year × Zone
7.4.1	RL1 — Baseline counts (per year)
7.4.2	RL1 — Longitudinal6 counts
7.4.3	RL1 — Longitudinal5 counts
7.4.4	RL2 — Baseline counts (per year)
7.4.5	RL2 — Longitudinal6 counts
7.4.6	RL2 — Longitudinal5 counts
7.4.7	RL3 — Baseline counts (per year)
7.4.8	RL3 — Longitudinal6 counts
7.4.9	RL3 — Longitudinal5 counts
7.4.10	Notes on overlapping membership across RL variants
7.5	Parcel-Size Mapping as a Cross-Check on the Most-Affected RL Variant
7.5.1	Objective and rationale
7.5.2	Inputs, selection, and residential filter
7.5.3	Quantile classification and symbology
7.5.4	What the map shows
7.5.5	Context from spot checks (qualitative, not used in quantitative inference)
7.5.6	Methodological notes and limitations

7.5.7	Implications for “most affected” RL selection
7.6	2021 Assessed Values Mapped by Decile (Citywide Patterns and RL Focus)
7.6.1	Purpose and inputs
7.6.2	Method (ArcGIS Pro workflow and symbology)
7.6.3	Citywide pattern in 2021
7.6.4	RL1 vs RL2: what the deciles add
7.6.5	Notes, limitations, and figure captions
7.7	Spatial Concentration of Black Residents (ACS Tract Choropleth)
7.7.1	Objective
7.7.2	Inputs and method
7.7.3	Cartographic decisions and what they imply
7.7.4	Read-out from the provided figure (descriptive, not inferential)
7.8	Northeast Lubbock after the 1970 Tornado: why much of the quadrant remains sparse or non-residential
7.8.1	Aim and scope
7.8.2	What happened on May 11, 1970 (path, intensity, and the northeast swath)
7.8.3	Immediate and medium-term rebuilding: urban renewal and civic investments
7.8.4	Today's land-use character: parkland, airport/industry corridors, limited housing pockets
7.8.5	How this context relates to the project's findings
7.9	Conclusions

List of Figures

Figure Number	Name
1.1	Archival microfilm frame of the Lubbock County Appraisal District Property Tax Roll, 1985
1.2	Value distribution (log10) — 1945 (ASSESSED)
1.3	Value distribution (log10) — 1975 (ASSESSED)
1.4	Value distribution (log10) — 1985 (ASSESSED)
1.5	Empirical CDF of values (log10) — 1945 vs 1975 vs 1985
1.6	Top 10 jurisdictions by record count — 1945 (Original Grantee City or Town)
1.7	Top 10 jurisdictions by record count — 1975 (Original Grantee City or Town)
1.8	Top 10 jurisdictions by record count — 1985 (CITY)
1.9	Exact uppercase address overlap across years
1.11	TotalValue distribution (log10) — 2012
1.12	TotalValue distribution (log10) — 2020
1.13	TotalValue distribution (log10) — 2021
1.14	Empirical CDF of TotalValue (log10) — 2012 vs 2020 vs 2021
1.15	LandValue vs ImpValue distributions (log10) — 2012
1.16	LandValue vs ImpValue distributions (log10) — 2020
1.17	LandValue vs ImpValue distributions (log10) — 2021
1.18	Top 10 neighborhoods by record count — 2012
1.19	Top 10 neighborhoods by record count — 2020
1.2	Top 10 neighborhoods by record count — 2021
1.21	Exact uppercase address overlap across 2012, 2020, and 2021
3.1	Possible redlining zone (per Adam Pirtle, Legal Aid of NorthWest Texas)
3.2	LBK6Y points over CityLimit, Council Districts and Redlining Zone
3.3	Core polygon and two buffer rings (schematic over basemap)
4.1	Bar chart — properties with data by year
4.2	Data availability by district and year
4.3	Data availability by zone and year
4.4	Heatmap of missingness by column and year
4.5	Small multiples — district×zone counts by year
4.6	Line chart of district averages by year
4.7	Grouped lines by zone
4.8	Waterfall panels for district growth intervals
5.1	PPFS box-and-whisker by zone 2021 (log y-axis)
6.1	Histograms (2025\$; All vs 281) — 1945
6.2	Histograms (2025\$; All vs 281) — 1975
6.3	Histograms (2025\$; All vs 281) — 1985
6.4	Histograms (2025\$; All vs 281) — 2012
6.5	Histograms (2025\$; All vs 281) — 2020

6.6	Histograms (2025\$; All vs 281) — 2021
6.7	Citywide PPFS (2025\$) Median by Year — 281 vs All
6.8	Citywide PPFS (2025\$) Median with IQR Ribbons
6.9	Boxplots (linear & log): 1945 L
6.1	Boxplots (linear & log): 1945 Log
6.11	Boxplots (linear & log): 1975 L
6.12	Boxplots (linear & log): 1975 Log
6.13	Boxplots (linear & log): 1985 L
6.14	Boxplots (linear & log): 1985 Log
6.15	Boxplots (linear & log): 2012 L
6.16	Boxplots (linear & log): 2012 Log
6.17	Boxplots (linear & log): 2020 L
6.18	Boxplots (linear & log): 2020 Log
6.19	Boxplots (linear & log): 2021 L
6.2	Boxplots (linear & log): 2021 Log
6.21	Zone medians 2021 (linear)
6.22	Zone medians 2021 (log)
6.23	Redlining Zone — trend of medians across years
6.24	Half-mile Buffer — trend of medians across years
6.25	Between 0.5–1 mile — trend of medians across years
6.26	Outside Both — trend of medians across years
6.27	Median log growth — 281 (within-address)
6.28	Median log growth — All (pseudo-panel)
6.29	F2025 Mean Value per SqFt - All vs Complete
6.3	F2025 Median Value per SqFt - By Zone (All)
6.31	F2025 Median Value per SqFt - By Zone (Complete)
6.32	F2025 Median Value per SqFt - 2021 by District (All)
6.33	F2025 Median \$/Sqft - %Change (2012→2021) by Zone (All)
6.34	Complete-to-All Median Ratio - 2021 by Zone
7.1	city_limit_1945, city_limit_1975, city_limit_1985, city_limit_2012, city_limit_2020, city_limit_2021
7.2	Lubbock ordinance context and road references
7.3	Possible redlining zones considered (RL1, RL2, RL3 and buffers)
7.4	All City Limits and Redlining Zones together
7.5	Baseline — Mean (in vs out) RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.
7.6	Baseline — Median (in vs out) RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.
7.7	Longitudinal6 — Mean (in vs out) Six-year balanced panel; RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.
7.8	Longitudinal6 — Median (in vs out) Six-year balanced panel; RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.
7.9	Longitudinal5 — Mean (in vs out) 1975, 1985, 2012, 2020, 2021; RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.

7.1	Longitudinal5 — Median (in vs out) 1975, 1985, 2012, 2020, 2021; RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.
7.11	Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 1945
7.12	Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 1975
7.13	Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 1985
7.14	Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 2012
7.15	Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 2020
7.16	Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 2021
7.17	Nominal Value (TotalValue) — Year 1945 — cohorts by zone×group
7.18	Nominal Value (TotalValue) — Year 1975 — cohorts by zone×group
7.19	Nominal Value (TotalValue) — Year 1985 — cohorts by zone×group
7.2	Nominal Value (TotalValue) — Year 2012 — cohorts by zone×group
7.21	Nominal Value (TotalValue) — Year 2020 — cohorts by zone×group
7.22	Nominal Value (TotalValue) — Year 2021 — cohorts by zone×group
7.23	Residential-scale parcels in city_limit_2021, colored by parcel-area deciles (dark = smallest; bright = largest), with RL1/RL2 outlines
7.24	Legend for deciles showing Shape_Area ranges (10 classes)
7.25	Parcels in city_limit_2021 symbolized by deciles of TotalValue_2021 (dark = higher, bright = lower), with RL1 and RL2 outlined
7.26	Legend showing the ten TotalValue_2021 class breaks used in FIG 7.25
7.27	Census tracts symbolized by Black population (ACS 5-year 2023), darker = higher count
7.28	Census tracts symbolized by White population (ACS 5-year 2023), darker = higher count
7.29	1970 Lubbock Tornado paths
7.30	Mackenzie Park & Canyon Lakes footprint

List of Tables

Table Number	Name
1.1	2012 Valuation Distribution Summary
1.2	2020 Valuation Distribution Summary
1.3	2021 Valuation Distribution Summary
2.1	1945 Address Pattern Library and Transformation Rules
2.2	Example of GeoPy House-Number Containment Acceptance
2.3	GeoPy Preprocessing and Query Context (Archival Years)
2.4	GeoPy Quality Assurance and Cleanup Rules
2.5	Audit and Canonical Column Lifecycle
2.6	Data Cleaning and Join Readiness Rules for Archival Data
4.1	Top 20 ZIP Codes by Property Record Count
6.1	Property Cohort Counts by Year
6.2	Zone Composition by Property Cohort
6.3	PPFS (2025\$) Summary Statistics by Year and Cohort
6.4	Nominal PPFS Medians by Year and Cohort
7.1	Mapping of Study Year to City Limit Boundary Decade
7.2	Comparison of Redlining Zone Variants by Median Value Gap (Baseline, Inflation-Adjusted)
7.3	RL1 Parcel Counts by Zone and Year — Baseline Cohort
7.4	RL1 Parcel Counts by Zone and Year — Longitudinal6 Cohort
7.5	RL1 Parcel Counts by Zone and Year — Longitudinal5 Cohort
7.6	RL2 Parcel Counts by Zone and Year — Baseline Cohort
7.7	RL2 Parcel Counts by Zone and Year — Longitudinal6 Cohort
7.8	RL2 Parcel Counts by Zone and Year — Longitudinal5 Cohort
7.9	RL3 Parcel Counts by Zone and Year — Baseline Cohort
7.1	RL3 Parcel Counts by Zone and Year — Longitudinal6 Cohort
7.11	RL3 Parcel Counts by Zone and Year — Longitudinal5 Cohort

Executive Summary

This report, **Lubbock Property Values & Redlining: A Chronological Report**, provides an in-depth analysis of the enduring effects of historical redlining on property values in Lubbock, Texas, from 1945 to 2021. The study focuses on three distinct historical redlining footprints, with **RL1—the Avenue C × 16th Street construction—identified as the operative footprint** due to its strong correlation with lower property values and higher Black population counts.

By analyzing archival and contemporary property valuation data, the report finds a clear association between historically redlined areas and present-day property values. The analysis of property value trends shows a consistent pattern: properties within the RL1 footprint have followed a distinctly different value trajectory compared to those outside this area.

The findings have significant implications for city leaders, urban planners, and policymakers. This evidence-based portrait of urban development can be used to inform targeted, equitable urban redevelopment strategies. While no direct causal claims are made, the descriptive alignment between historical redlining maps, demographic data, and property values serves as a crucial resource for understanding and addressing the long-term legacy of discriminatory housing policies in Lubbock.

Project Objective

This project examines the trajectory of property values in Lubbock, Texas, over the period from 1945 to 2021, with particular attention to the long-term impact of historical redlining practices. Each stage of the study will be documented in the precise order it was undertaken, ensuring the report reflects both the methodological flow and the historical unfolding of the research.

Introduction

Across the United States, discriminatory housing and lending practices of the twentieth century created durable spatial patterns of advantage and disadvantage. Although federal redlining—most notably associated with the Home Owners’ Loan Corporation (HOLC) maps—was outlawed decades ago, a substantial body of research shows that the neighborhoods targeted by such practices continue to experience lower property values, reduced homeownership, and diminished access to credit and amenities. In West Texas, Lubbock’s built environment and planning history reflect many of these national dynamics through a combination of racially segregated settlement patterns, mid-century urban renewal, and later zoning and infrastructure decisions that concentrated environmental burdens and disinvestment in specific areas, particularly in and around East Lubbock.

This study is motivated by the need to document how those historical decisions may be reflected in long-run property value trajectories at the parcel level. Rather than relying on a single snapshot, it follows the same addresses across multiple benchmark years between 1945 and 2021. By tracing these longitudinal patterns and comparing them across regions within the city, the project seeks to clarify whether neighborhoods associated with historic segregation and disinvestment have followed distinct price paths relative to other parts of Lubbock. The emphasis on common addresses and consistent temporal checkpoints is intended to reduce cross-sectional bias and to foreground change over time.

The results are designed to be decision-supporting. Local policymakers and planners can use the findings to ground discussions of equitable urban redevelopment, including where to target infrastructure upgrades, housing rehabilitation, environmental remediation, and incentives that close historical gaps in neighborhood investment. By offering a clear, evidence-based portrait of how values evolved in different parts of Lubbock—and how those trends intersect with the city’s planning legacy—this report aims to inform practical choices in zoning updates, corridor reinvestment, and neighborhood revitalization strategies. Ultimately, the study provides a historical baseline and a comparative lens that can help align future development with goals of inclusion, resilience, and long-term economic vitality for all residents.

Chronological Development

The narrative will begin with the earliest stage of the work, describing how data was first located, prepared, and analyzed. As the story progresses, each subsequent phase will build on the previous one, reflecting both the logical and temporal sequence of the project. Within each stage, the discussion will present the context and purpose of the step, the specific materials and methods used, and the findings derived from your analyses. Any limitations or challenges encountered will be discussed in context, without introducing interpretations or figures beyond the study.

1. Data Origin and Collection (Archival Years: 1945, 1975, 1985)

The earliest phases of the study relied on archival images of the Lubbock County Appraisal District's property tax rolls. These materials were preserved on computer-output microfilm and viewed on a microfilm reader; the frames display fixed-width records for individual parcels along with page-level identifiers. A representative image (see figure 1.1) from 1985 shows the header "PROPERTY TAX ROLL FOR 1985," a printed date of 09/23/85, and device branding ("Datagraphix"), with page references such as "Selector = 13" and sequential page numbers. Each boxed record on the frame corresponds to a single parcel entry and presents the situs address, legal description, map sheet reference, internal control codes, assessed components, exemptions, and the tax levies attributed to specific jurisdictions.

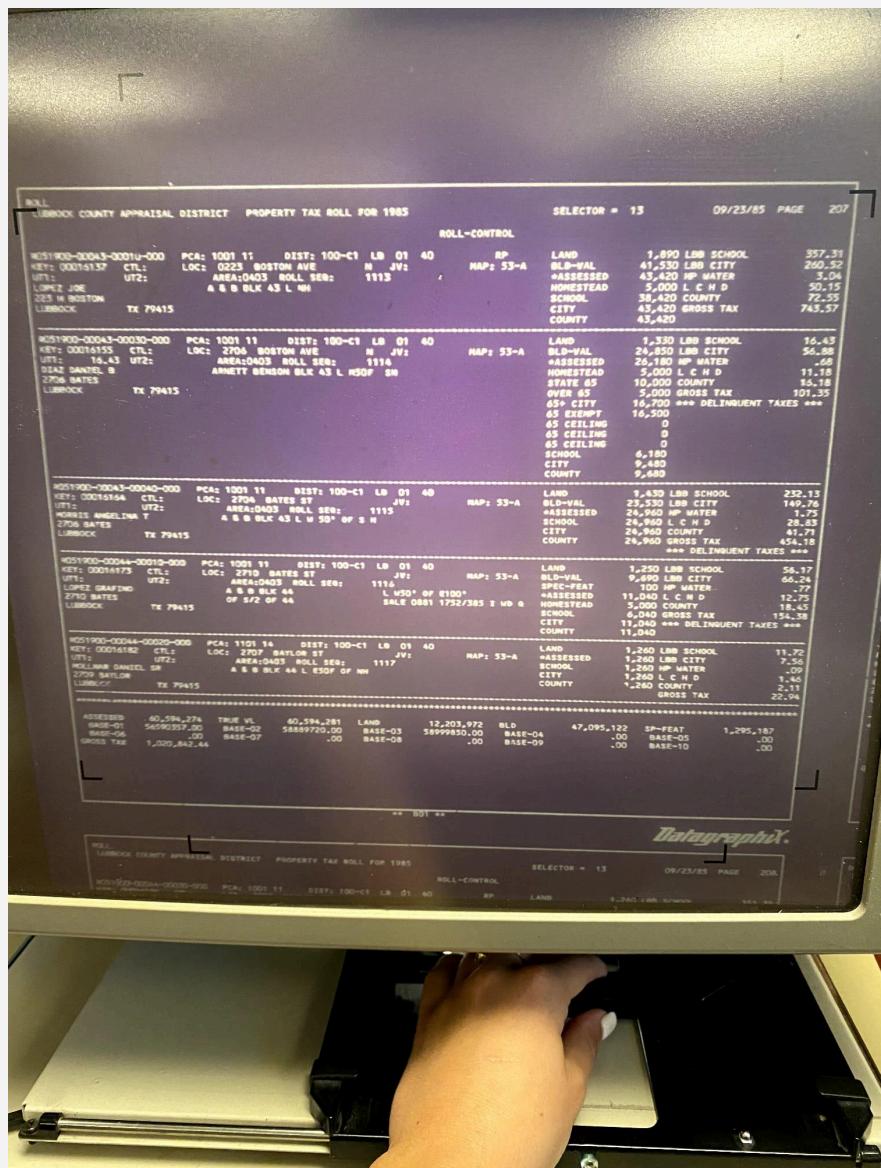


Figure 1 — Archival microfilm frame of the Lubbock County Appraisal District Property Tax Roll, 1985 (Datagraphix COM; Selector 13, 09/23/85)

From these frames, the study assembled the 1945, 1975, and 1985 data through manual transcription. Researchers reviewed the microfilm images on the reader, located relevant parcel entries, and keyed the information directly into spreadsheet workbooks. Optical character recognition was not used at this stage, given the quality and formatting of the archival images; the objective was to preserve the original field structure and abbreviations as printed on the rolls. The resulting spreadsheets therefore mirror the source formatting while rendering the information machine-readable for downstream analysis.

The transcribed fields include the printed situs address (as shown on the roll), the legal description (block, lot, and subdivision references), map sheet identifiers (for example, codes like “MAP: 53-A”), and internal roll controls (such as PCA and DIST codes where present). Value fields were captured as printed, including land value, building value, and the assessed total. Where the rolls enumerate exemptions or flags, such as homestead entries, senior caps, or delinquent notices, those indicators were recorded verbatim. Jurisdictional levy lines—commonly listing school, city, county, and special districts (for example, water or hospital districts)—were entered as separate fields to retain the structure of the original ledger.

To maintain traceability, each transcribed parcel line retains a citation to its source frame, including the printed page number and, when visible, the selector number and frame date. This provenance allows any value or address in the spreadsheets to be checked against the original archival image. Because manual transcription can introduce typographical error, the study preserved the original abbreviations and punctuation from the rolls and deferred any normalization of addresses or legal descriptions to later, explicitly documented preprocessing steps.

This manual collection phase establishes the foundation for the longitudinal analysis by creating a consistent, parcel-level record for the archival years. Subsequent sections of the report (developed once the spreadsheet workbooks are introduced) will describe how these transcribed fields were standardized, matched across time, and integrated with later-year datasets to enable address-level comparisons between 1945 and 2021.

1.1 Archival Datasets — Description & Exploratory Analysis (1945, 1975, 1985)

This section extends the *Data Origin and Collection* description by characterizing the three archival spreadsheets in quantitative detail and by presenting an initial exploratory analysis. All numerics reported here are computed directly from the initial datasets of years 1945, 1975 and 1985, using only fields present in those files and the explicitly constructed variable noted below.

1.1.1 Dataset characterization

The archival spreadsheets correspond to the years **1945**, **1975**, and **1985**. The 1945 file contains **10,653** rows and **18** columns; the 1975 file contains **61,601** rows and **26** columns; and the 1985 file contains **25,978** rows and **21** columns. In each file, the working address field is the column labeled **ADDRESS**, used here strictly as provided (upper-cased for comparison; no normalization beyond casing).

Because the microfilm ledger prints multi-line entries and repeated levy rows, duplicate address strings are expected. Counting exact uppercase strings in **ADDRESS** yields:

- **1945:** 3,927 distinct addresses (36.86% of rows), 7,037 rows participating in duplicates (66.06%), and 0 blanks.
- **1975:** 37,937 distinct addresses (61.59% of rows), 30,389 duplicate rows (49.33%), and 1 blank.
- **1985:** 18,689 distinct addresses (71.94% of rows), 10,415 duplicate rows (40.09%), and 0 blanks.

Cross-year continuity using exact uppercase string matches is limited over longer horizons: 3,035 common address strings between **1975** and **1985** (\approx 8.00% of 1975 distinct; \approx 16.24% of 1985 distinct), 9 between **1945** and **1975** (\approx 0.23% of 1945 distinct), 6 between **1945** and **1985** (\approx 0.15% of 1945 distinct), and 4 shared across **all three** years. These counts motivate the address standardization and legal-description crosswalks introduced later.

1.1.2 Valuation fields used in EDA

For 1945, a working valuation variable **ASSESSED 1945** is constructed **row-wise as the maximum of Value of City Property and Value of Personal Property** (both fields are present in the file). For 1975, the column **Value of City Property (Total County Value)** is used as provided. For 1985, the column **ASSESSED** is used as provided. All values are treated as nominal dollars exactly as stored; no deflation or re-indexing is applied in this stage.

Coverage and summary quantiles for these year-specific valuation measures are:

- **1945 — ASSESSED 1945:** non-null positive count 7,745 (72.70% of rows). Quantiles (USD): p10 60, p25 250, median 800, p75 1,800, p90 2,624, p95 4,000, p99 15,856; min positive 1; max 520,350. Interquartile range (IQR): 1,550.
- **1975 — Value of City Property (Total County Value):** non-null positive count 61,567 (99.94% of rows). Quantiles (USD): p10 240, p25 1,000, median 2,560, p75 5,250, p90 8,700, p95 11,760, p99 40,083; min positive 10; max 2,520,810. IQR: 4,250.
- **1985 — ASSESSED:** non-null positive count 25,858 (99.54% of rows). Quantiles (USD): p10 2,810.7, p25 9,650, median 22,363, p75 44,602.25, p90 64,473.6, p95 94,836.6, p99 344,297.5; min positive 10; max 9,062,880. IQR: 34,952.25.

These distributions are heavy-tailed and right-skewed in each year. To summarize them without undue leverage from outliers, the figures below use **log10 transforms** (i.e., $\log_{10}(\text{value in USD})$) for histograms and plot **empirical cumulative distribution functions (ECDFs)** for cross-year comparisons.

1.1.3 Regional coverage indicators

Regional or jurisdictional indicators vary by year. For **1945** and **1975**, **Original Grantee City or Town** provides a coarse jurisdiction label; for **1985**, a **CITY** field appears with predominantly numeric codes. Record concentration in the most common categories is as follows (top entries with shares of total rows):

- **1945 (Original Grantee City or Town):** NAN 5,222 (49.02%) [missing/unspecified], LUB 1,536 (14.41%), " 729 (6.84%), SLATON 305 (2.86%), NE/4 223 (2.09%).
- **1975 (Original Grantee City or Town):** LUBBOCK 46,277 (75.12%), NAN 5,351 (8.69%), LUBBOCK TX 3,158 (5.13%), IDALOU 820 (1.33%), SHALLOWATER 616 (1.00%), followed by WOLFFORTH 378, ABERNATHY 305, DALLAS 254, HOUSTON 207, AMARILLO 141.

- 1985 (CITY): NAN 775 (2.98%), 0 690 (2.66%), 4650 87 (0.33%), 4400 81 (0.31%), 650 66 (0.25%), with additional codes in smaller counts. The prevalence of missing or numeric codes indicates the need for a codebook or crosswalk prior to regional aggregation.
-

1.1.4 Figures and interpretations

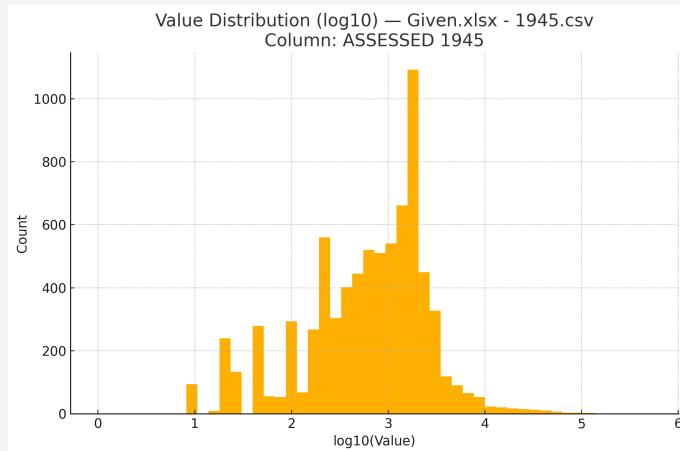


Figure 1.2 — Value distribution (log10) — 1945 (ASSESSED)

The 1945 distribution is highly right-skewed with a dense mass below \$2,000 and an extended upper tail. The **median** is \$800, and 90% of positive observations lie below \$2,624; the **99th percentile** is \$15,856, while the **maximum** reaches \$520,350, confirming the presence of a small number of very high-value entries. The **IQR** of \$1,550 (p25–p75) indicates tight central dispersion relative to later years.

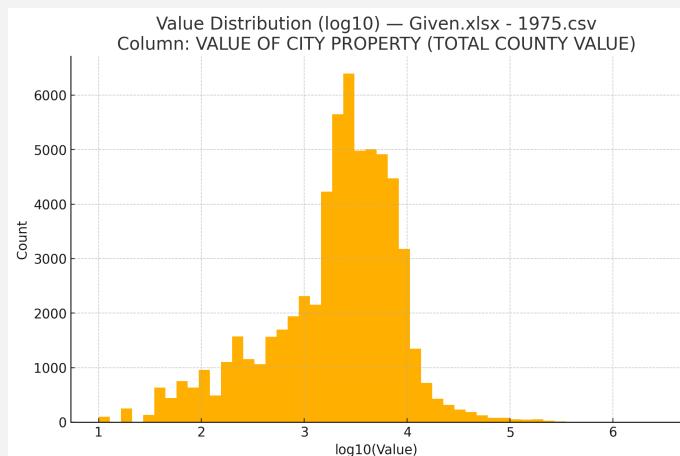


Figure 1.3 — Value distribution (log10) — 1975 (ASSESSED)

The 1975 distribution shifts materially to the right relative to 1945. The **median** is \$2,560 with p75 at \$5,250 and p90 at \$8,700, implying that 90% of parcels are valued below \$8.7k on this measure. The **99th percentile** at \$40,083 and **maximum** \$2,520,810 indicate a heavier upper tail than 1945. Coverage is nearly complete (99.94% non-null positives), so distributional inferences are robust to missingness in this field.

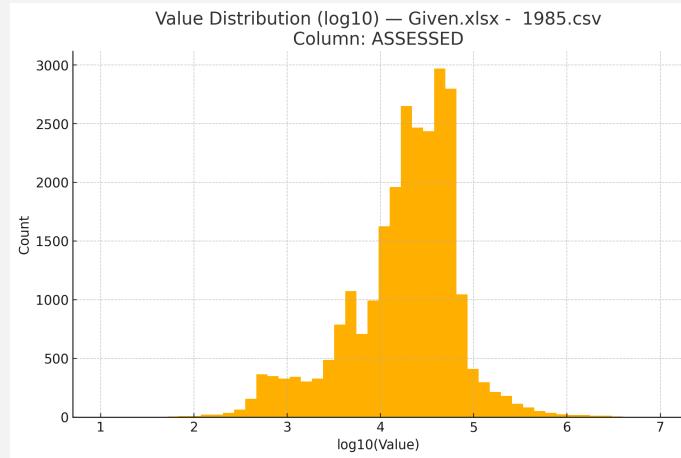


Figure 1.4 — Value distribution (log10) — 1985 (ASSESSED)

The 1985 distribution exhibits a substantial rightward shift and widening dispersion. The **median** is **\$22,363**, with **p75** at **\$44,602**, **p90** at **\$64,474**, and **p95** at **\$94,837**. The **99th percentile** is **\$344,298**, and the **maximum** is **\$9,062,880**, producing a very long tail characteristic of urban appraisal rolls with a small number of large commercial or institutional parcels. The **IQR** of **\$34,952** underscores a markedly higher mid-spread than in 1975.

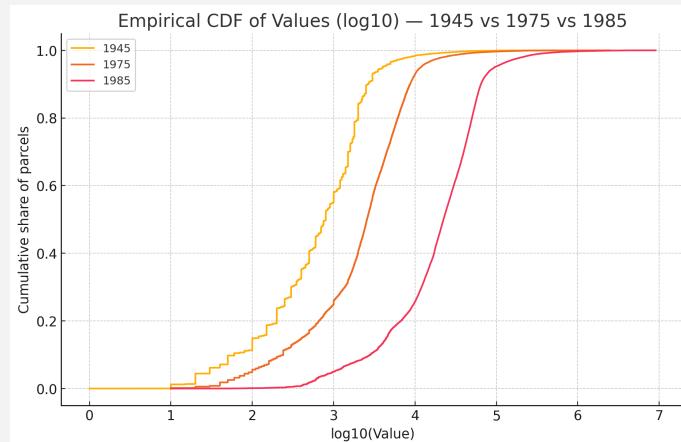


Figure 1.5 — Empirical CDF of values (log10) — 1945 vs 1975 vs 1985

The ECDF confirms first-order stochastic dominance moving from 1945 → 1975 → 1985 on the year-specific valuation measures. At the **50th percentile**, the value rises from **\$800** (1945) to **\$2,560** (1975) to **\$22,363** (1985). At the **90th percentile**, the trajectory is **\$2,624** → **\$8,700** → **\$64,474**. These gaps quantify the broad right-shift in the valuation distribution over time, while the separation between curves at upper quantiles evidences a thickening upper tail by 1985.

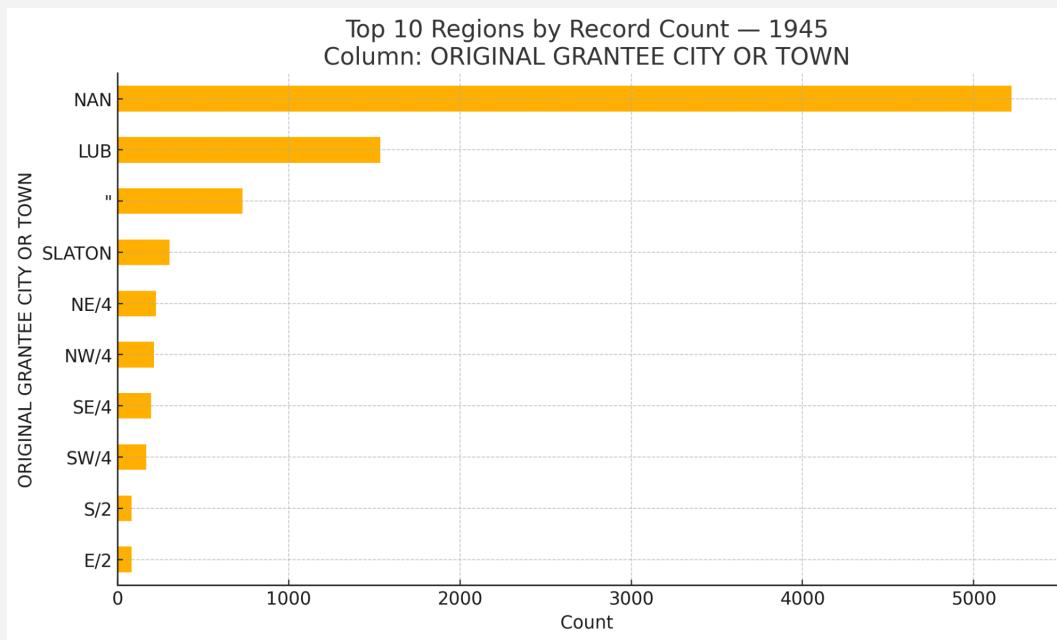


Figure 1.6 — Top 10 jurisdictions by record count — 1945 (Original Grantee City or Town)

Jurisdictional coverage is dominated by missing/unspecified entries ([NAN 49.02%](#)) and a large [LUB](#) class ([14.41%](#)). Named places such as [SLATON \(2.86%\)](#) appear in much smaller proportions. The presence of labels like [NE/4](#), [NW/4](#), [SE/4](#) suggests quarter-section descriptors rather than municipal names, indicating heterogeneous semantics in this field for 1945. Any regional aggregation will require harmonization.

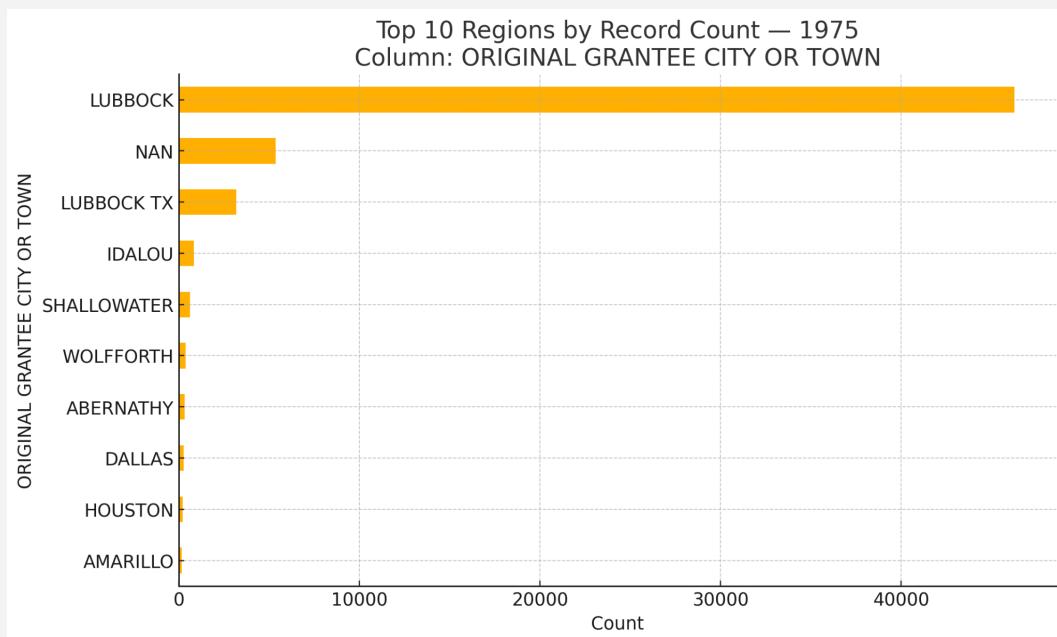


Figure 1.7 — Top 10 jurisdictions by record count — 1975 (Original Grantee City or Town)

Coverage is concentrated in LUBBOCK (46,277 rows; 75.12%), followed by NAN (8.69%) and LUBBOCK TX (5.13%). Surrounding municipalities (e.g., IDALOU 1.33%, SHALLOWATER 1.00%, WOLFFORTH 0.61%) account for modest shares. The dominance of LUBBOCK aligns with expectations for the urban core and provides a strong basis for within-city analysis in this year.

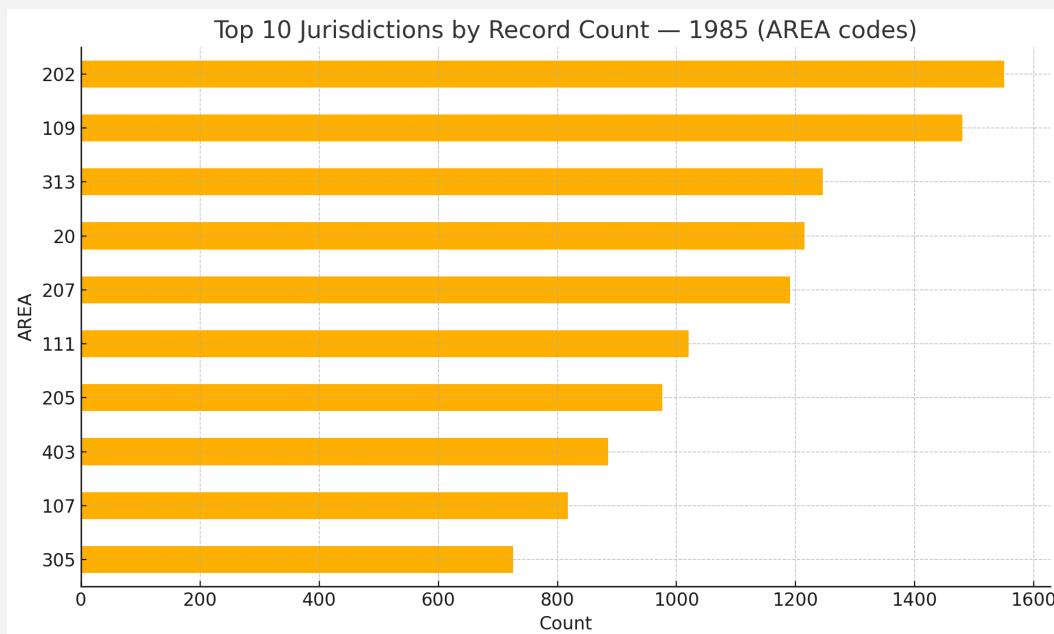


Figure 1.8 — Top 10 jurisdictions by record count — 1985 (CITY)

The 1985 AREA field is a coded jurisdiction indicator rather than a named place label. Record counts are dispersed across many codes, with no dominant category: the top entries are 202 (1,550; 5.97%), 109 (1,480; 5.70%), 313 (1,246; 4.80%), 20 (1,215; 4.68%), and 207 (1,191; 4.58%), followed by 111 (1,020; 3.93%), 205 (976; 3.76%), 403 (885; 3.41%), 107 (818; 3.15%), and 305 (726; 2.79%). Collectively, these top ten codes account for ≈42.8% of all 1985 rows, indicating broad distribution of parcels across many appraisal areas rather than concentration in a few. Because AREA is numeric and unlabeled in the dataset, a codebook or crosswalk is required before drawing spatial conclusions or aligning these areas with historical redlining or later regional definitions.

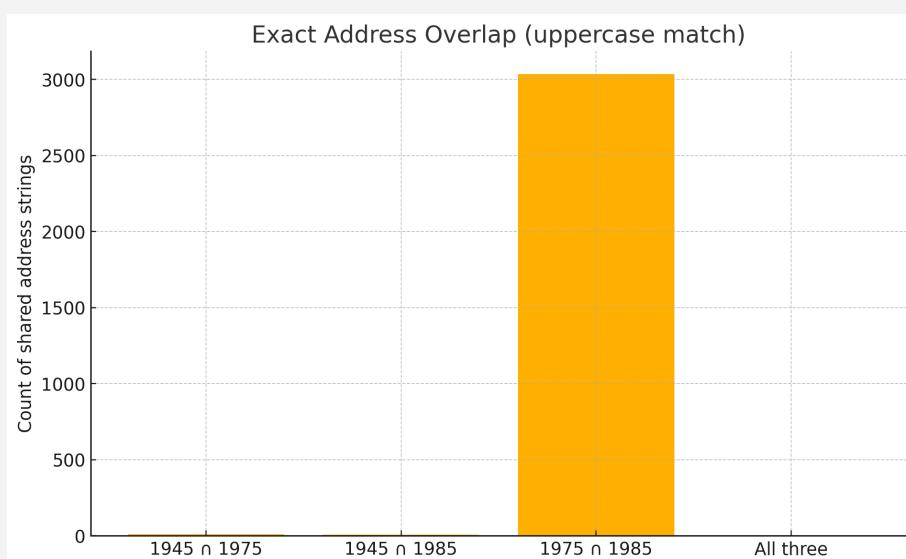


Figure 1.9 — Exact uppercase address overlap across years

Overlap is **3,035** addresses for **1975** \cap **1985**, **9** for **1945** \cap **1975**, **6** for **1945** \cap **1985**, and **4** across **all three**. Relative to distinct counts, the **1975** \cap **1985** overlap equals $\approx 8.00\%$ of 1975 distinct and $\approx 16.24\%$ of 1985 distinct, while 1945 overlaps are near zero. These diagnostics substantiate the need for deterministic or probabilistic address standardization keyed to legal descriptions for longitudinal linkage.

Notes. All statistics above are computed from the provided files in nominal dollars. Log-scaled axes (\log_{10}) are used for distribution plots; percentages are row-share calculations within each year's file. No external datasets or invented metrics are introduced in this section.

1.2 Contemporary Datasets — Provenance & Exploratory Analysis (2012, 2020, 2021)

The contemporary phase draws on three standardized parcel extracts obtained from **public internet** sources: **2012**, **2020**, and **2021**. Unlike the archival microfilm transcriptions, these files include rich, machine-readable attributes for situs address, neighborhood codes, valuation components, and structure characteristics. All statistics below are computed directly from the provided CSVs without external augmentation or invented constructs.

1.2.1 File structure and address diagnostics

The 2012 file contains 111,011 rows; the 2020 file 122,479 rows; and the 2021 file 124,779 rows. Using the **Situs** field (upper-cased, no normalization beyond casing) as the working address key yields the following:

- **2012:** **101,586** distinct addresses (**91.51%** of rows); **10,815** rows participate in duplicate address clusters (**9.74%**); **0** blanks.
- **2020:** **113,042** distinct (**92.30%**); **10,937** duplicate rows (**8.93%**); **0** blanks.
- **2021:** **115,334** distinct (**92.43%**); **10,968** duplicate rows (**8.79%**); **0** blanks.

Exact, upper-case address overlaps indicate very strong continuity across contemporary years: **98,869** common addresses for **2012** \cap **2020**, **98,549** for **2012** \cap **2021**, **112,626** for **2020** \cap **2021**, and **98,537** that appear in **all three** files. Expressed as set coverage, **97.33%** of 2012 addresses reappear in 2020, **97.01%** reappear in 2021, and **99.63%** of 2020 addresses reappear in 2021. These counts establish a robust base for longitudinal address-level comparisons within the 2012–2021 window.

1.2.2 Valuation components and internal consistency

Each file provides three valuation fields—**TotalValue**, **LandValue**, and **ImpValue** (improvements). A file-wide arithmetic check confirms that, where present on a row, **TotalValue = LandValue + ImpValue**; the median difference is **0** in **2012** ($n=111,011$), **2020** ($n=122,479$), and **2021** ($n=124,779$). This identity enables like-for-like distributional comparisons over time.

1.2.3 Distributional summaries (nominal USD)

The following tables report percentiles for `TotalValue`, `LandValue`, and `ImpValue` exactly as stored (nominal USD). Counts refer to the number of **positive** observations used in the percentile calculations.

Field	Count (pos.)	p10	p25	Median	p75	p90	p95	p99	Max
TotalValue	110,095	9,600	35,170.50	84,269	134,385	222,037	320,432	919,419.20	118,091,610
LandValue	110,978	0	0	8,000	20,000	35,152.90	50,000	183,935.18	46,011,326
ImpValue	110,672	2,000	31,000	89,513	139,717	226,153.40	321,160.60	929,651.09	118,091,610

Table 1.1 — 2012 Valuation Distribution Summary

Field	Count (pos.)	p10	p25	Median	p75	p90	p95	p99	Max
TotalValue	120,727	14,553.80	52,135.50	123,349	190,002	314,395.40	447,581.30	1,396,449.22	135,523,505
LandValue	121,737	100	8,500	14,300	28,000	48,358.30	69,000	227,632.15	15,382,531
ImpValue	120,934	1,950	43,000	121,323	188,581	314,364.10	444,938.40	1,391,647.37	135,523,505

Table 1.2 — 2020 Valuation Distribution Summary

Field	Count (pos.)	p10	p25	Median	p75	p90	p95	p99	Max
TotalValue	123,246	15,120	59,073.25	136,740	208,745.75	341,081	478,770.75	1,383,280.65	135,523,505
LandValue	124,020	100	10,000	14,750	30,000	52,000	74,000	258,306	22,000,000
ImpValue	123,339	2,400	48,000	136,185	212,064	339,629.10	443,927.40	1,242,541.27	135,523,505

Table 1.3 — 2021 Valuation Distribution Summary

1.2.4 Structure age (YearBuilt)

`YearBuilt` appears in all three files. Reported percentiles (including printed zeros where present) are:

- **2012**: p25 **1945**, median **1965**, p75 **1987**, p90 **2003** (min **0**).
- **2020**: p25 **1947**, median **1971**, p75 **1997**, p90 **2010** (min **0**).
- **2021**: p25 **1941**, median **1967**, p75 **1997**, p90 **2012** (min **0**).

Zeros likely encode missing construction years and will be handled explicitly in the linkage stage; they are left as-is here to reflect the raw files.

1.2.5 Neighborhood coverage (NbhdCode)

Neighborhood codes (`NbhdCode`) are populated at high rates in all three years. The most frequent codes account individually for ~2% of records, indicating broad dispersal across many neighborhoods. For example, the top codes are:

- **2012**: **1101 2,517 (2.27%)**, **1100 2,265 (2.04%)**, **0401NORTH2 1,852 (1.67%)**.

- 2020 : 1100 2,343 (1.91%), 1101 2,167 (1.77%), 0401NORTH2 1,915 (1.56%).
 - 2021: 1100 2,275 (1.82%), 1101 2,253 (1.81%), 0401NORTH2 1,906 (1.53%).
 - These frequencies facilitate regionally-stratified comparisons with historic patterns once a crosswalk from codes to geography is applied.
-

1.2.6 Figures and interpretations (contemporary years)

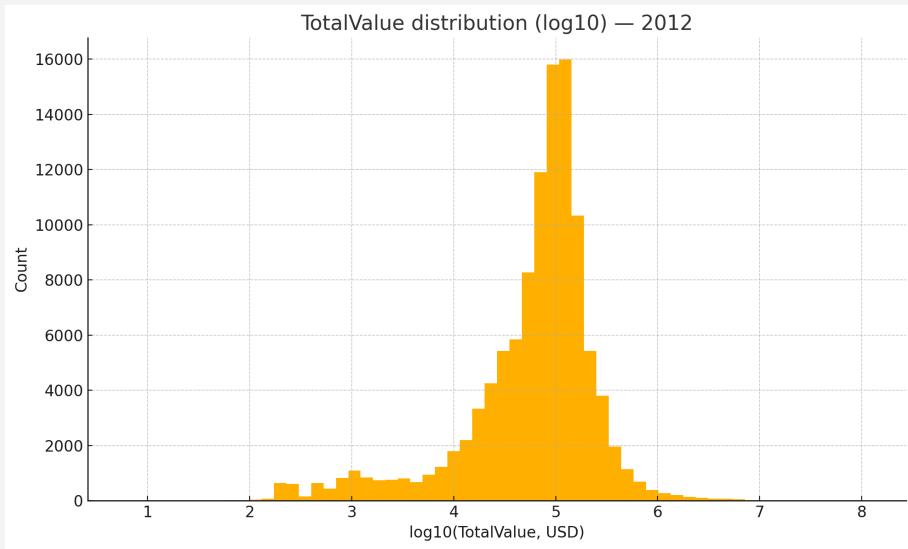


Figure 1.11 — TotalValue distribution (log10) — 2012

The mass of the distribution sits around $\text{log10} \approx 4.9\text{--}5.2$ ($\approx \$80\text{k}\text{--}\160k), with a long right tail reaching beyond **\$100M**. The **median** (**\$84,269**) and **IQR** (**\$99,214.5**) indicate moderate dispersion relative to later years.

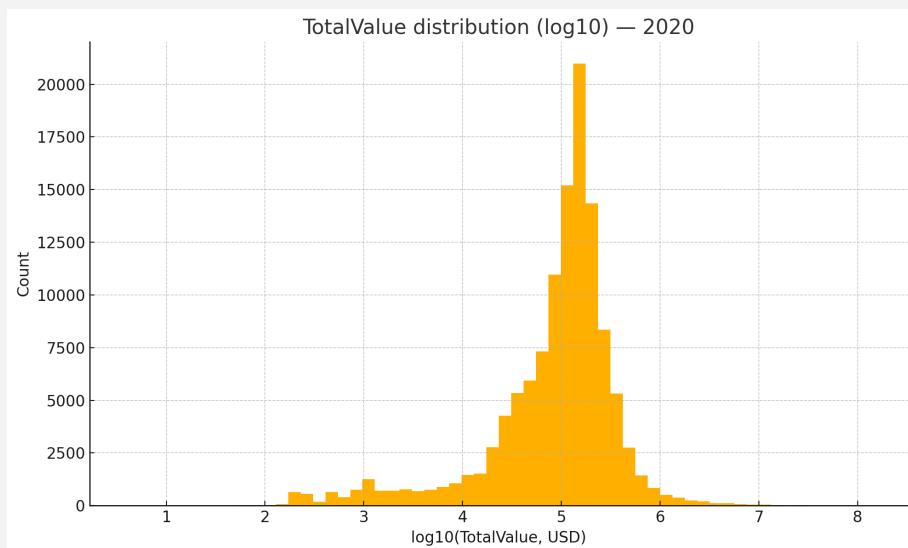


Figure 1.12 — TotalValue distribution (log10) — 2020

The distribution shifts right relative to 2012 with heavier upper quantiles (median **\$123,349**; p90 **\$314,395**). The IQR widens to **\$137,866.5**, signaling broader mid-spread consistent with market appreciation and/or reassessment changes between 2012 and 2020.

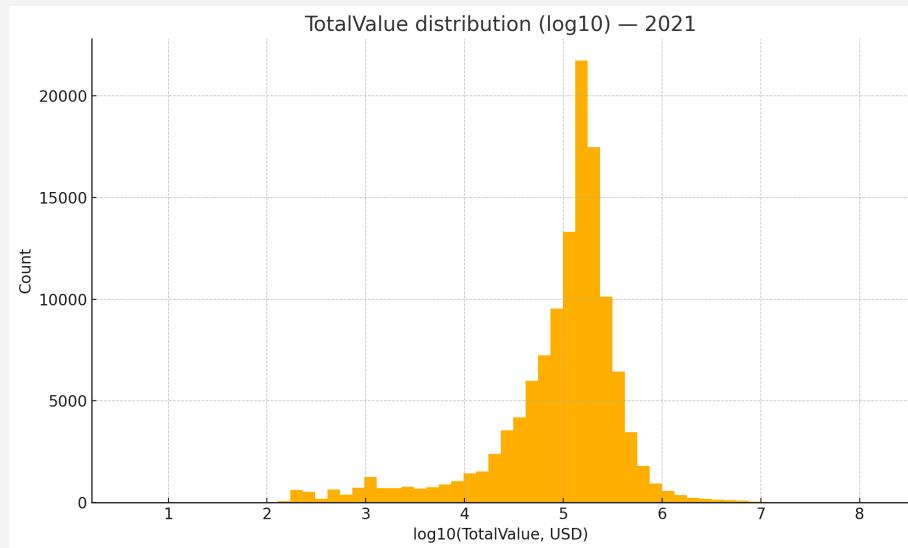


Figure 1.13 — TotalValue distribution (log10) — 2021

Another rightward nudge is visible (median **\$136,740**; p90 **\$341,081**), with the upper tail close to 2020's extent (p99 $\approx \$1.38M$). The mid-spread expands further (**IQR \$149,672.5**), maintaining a right skew.

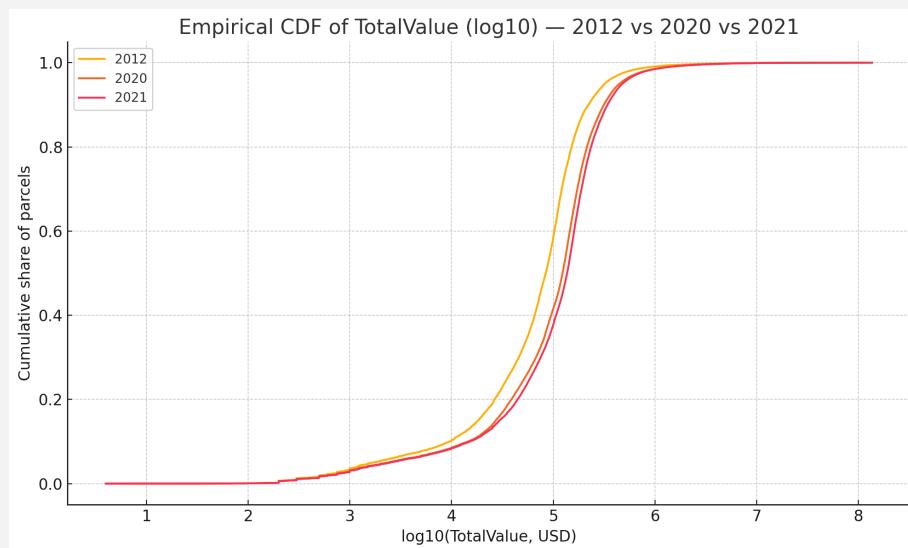


Figure 1.14 — Empirical CDF of TotalValue (log10) — 2012 vs 2020 vs 2021

The ECDF displays first-order dominance from **2012 → 2020 → 2021**: at the **median**, values rise **\$84k → \$123k → \$137k**, and at **p90**, **\$222k → \$314k → \$341k**. Curve separation is greatest between 2012 and 2020 and smaller between 2020 and 2021, indicating deceleration at the upper quantiles in the final year.

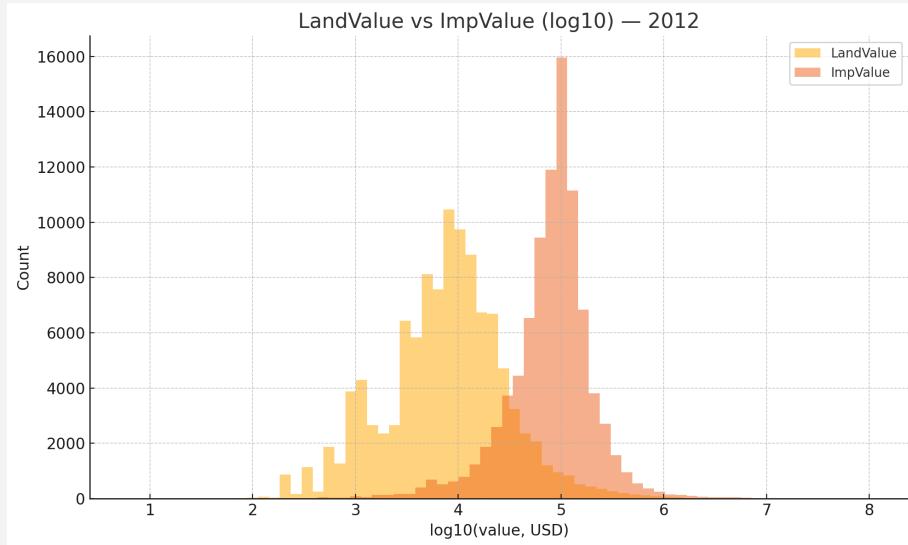


Figure 1.15 — LandValue vs ImpValue distributions (log10) — 2012

Improvements dominate assessed value, with **median ImpValue** \$89,513 versus **median LandValue** \$8,000. The distributions are well separated on the log scale, and the tail of ImpValue extends into the multi-million range.

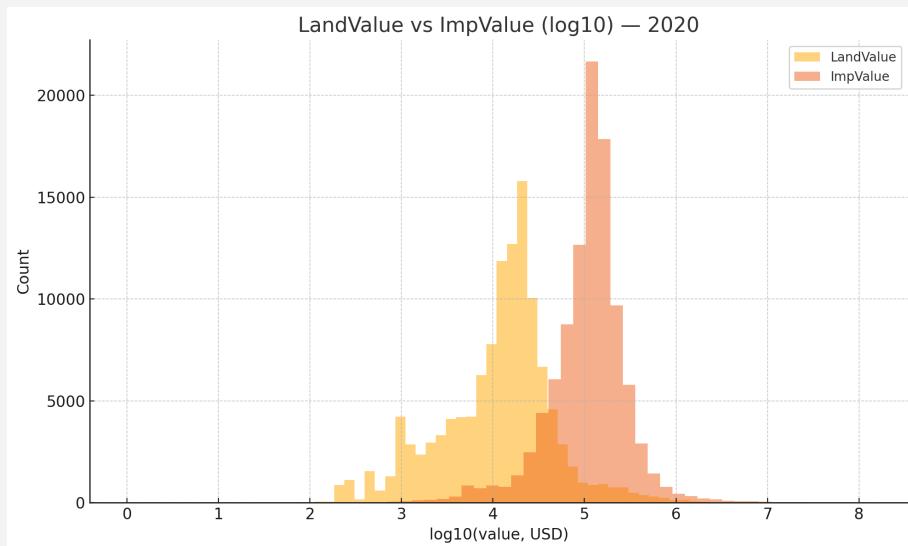


Figure 1.16 — LandValue vs ImpValue distributions (log10) — 2020

Both components shift right; **median land** rises to \$14,300 and **median improvements** to \$121,323. The widening gap mirrors the total-value expansion documented in figure 1.12.

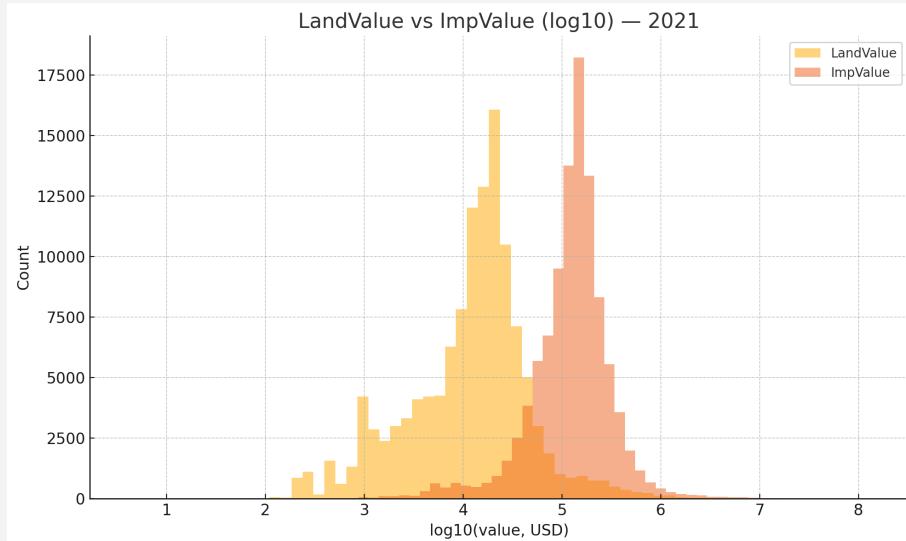


Figure 1.17 — LandValue vs ImpValue distributions (log10) — 2021

The right-shift continues (land \$14,750; improvements \$136,185), with a thicker upper tail for improvements ($p95 \approx \$444k$; $p99 \approx \$1.24M$), consistent with capital-intensive properties driving the extreme right tail of [TotalValue](#).

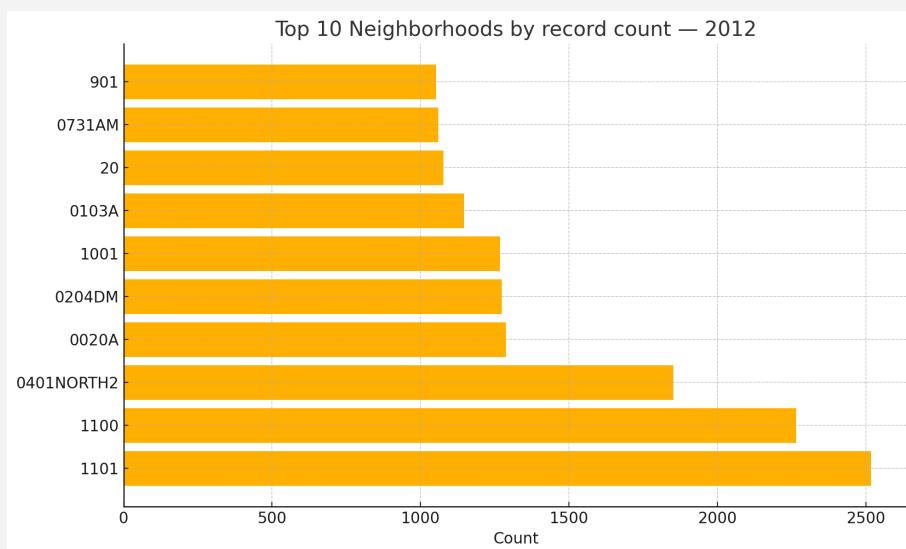


Figure 1.18 — Top 10 neighborhoods by record count — 2012

No single neighborhood dominates; the leading codes (1101, 1100, 0401NORTH2) account individually for $\approx 2\%$ of rows. This dispersion supports neighborhood-level stratification without singular hotspots overwhelmingly driving totals.

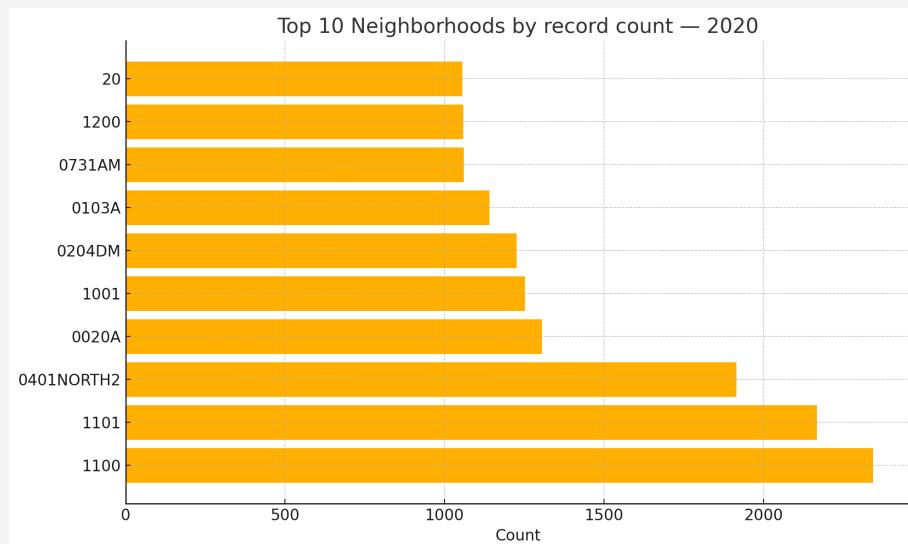


Figure 1.19 — Top 10 neighborhoods by record count — 2020

The ranking remains stable (**1100**, **1101**, **0401NORTH2**), with shares between ~1.5–1.9% each, indicating continuity in the geographic distribution of records.

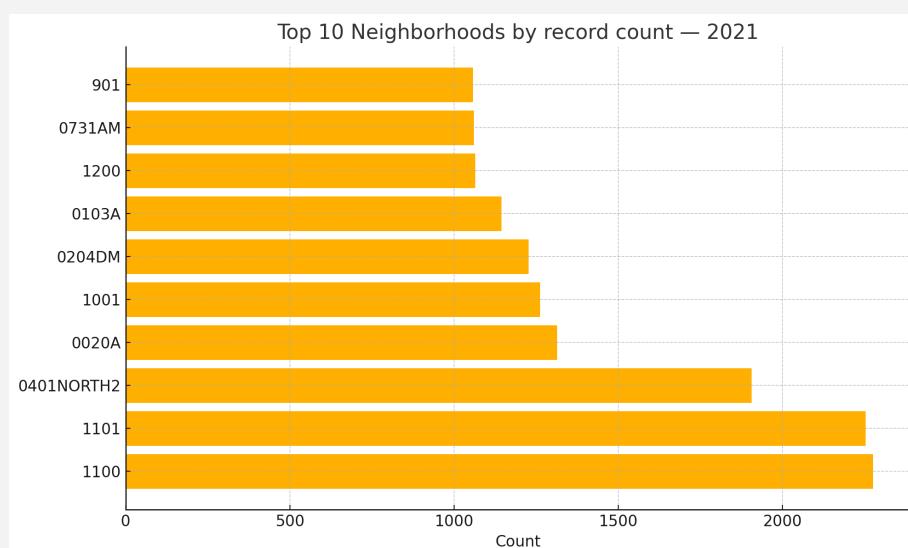


Figure 1.20 — Top 10 neighborhoods by record count — 2021

The top codes persist (**1100**, **1101**, **0401NORTH2**) with similar shares (~1.5-1.8%), maintaining the even spread across neighborhoods seen in earlier years.

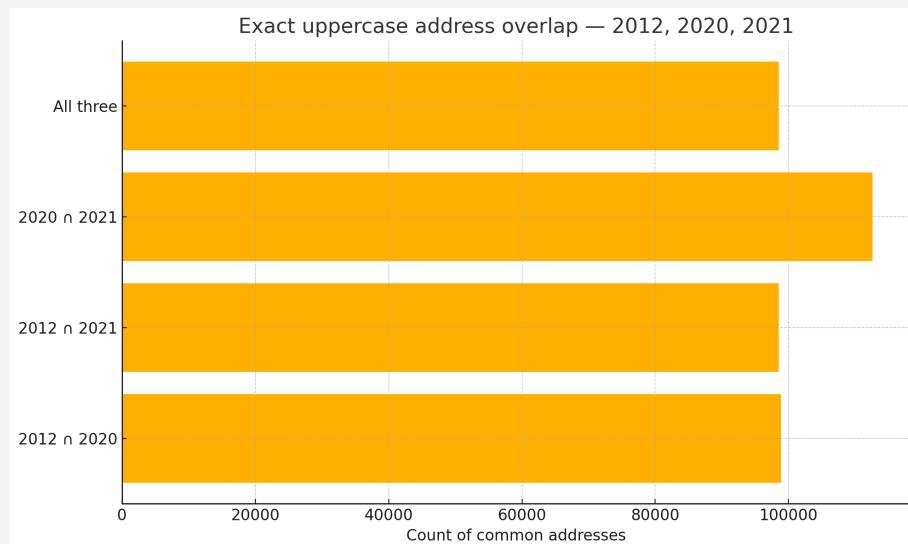


Figure 1.21 — Exact uppercase address overlap across 2012, 2020, and 2021

The overlap counts (**98,869**, **98,549**, **112,626**; all three **98,537**) confirm that the vast majority of addresses persist across years, simplifying linkage and enabling within-address growth-rate calculations in the next stage of the study.

Notes. All values are reported in nominal dollars. Plots use log10 scales for value distributions to reduce leverage from extreme upper-tail observations; percentages are computed as row shares within each file. Zeros in `YearBuilt` are preserved as recorded and treated as missing codes in subsequent standardization.

2. Methods: Data Cleaning, Address Standardization & Merging

2.0 Overview of the pipeline

The standardization approach evolved from rule-based tokenization to geocoding-backed canonicalization. Three conceptual iterations of standardization has been done to the addresses:

- **Iteration 1 — Regex/Token Standardization:** Bring heterogeneous 1945 strings into a machine-parsable form; apply light normalization to 1975/1985; create per-year keyword strings and split them into token families used by a vectorized diagnostic matcher.
 - **Iteration 2 — GeoPy-based normalization:** Use a geocoder to obtain more consistent address components while retaining original text and Iteration-1 outputs.
 - **Iteration 3 — Google Geocoding API canonicalization:** Submit cleaned strings to Google, capture structured components/coordinates, and define the final canonical address key for linkage and mapping.
-

2.1 Iteration 1 — Regex & Token Standardization

This process is to convert irregular 1945 address strings into a consistent **temporary canonical text** for downstream matching and apply light normalization to 1975/1985 (because they are already more regular). Also to build token families to enable fast, vectorized screening of cross-year candidates.

keywords_1945 / keywords_1975 / keywords_1985 — These are intermediate canonical strings per year. These columns are created in the respective dataset and each word, numericals or words, excluding special characters are extracted from the addresses and they are stored in these new columns in lists. They are then categorized as different token families mentioned below.

- **Token families per year:**
 - **NK:** numeric tokens (e.g., house numbers, numbered streets).
 - **ASK:** short alpha tokens (length ≤ 2 ; e.g., N, E, NE, SE).
 - **ALK:** long alpha tokens (> 2 chars; e.g., BROADWAY, AVENUE).
- **Audit columns:** original row index (e.g., `oldindex_1945`) and original address text preserved.

2.1.1 1945 workflow (irregular strings)

1. **Load & audit preservation.** The original row index is copied to an audit column to withstand row moves and filtering.
2. **Punctuation/sentinel filtering.** Pure punctuation placeholders (e.g., ' ', " ", and repeated variants) are treated as invalid and excluded. No additional sentinels (e.g., the words *same/ditto*) are used at this stage.

3. **Small token fixes.** Selected short forms are expanded or normalized (e.g., LUB → LUBBOCK, BX → BOX).
4. **Ordinal spacing fix.** Numeric+ordinal strings are normalized (e.g., 23RD → 23 RD).
5. **Pattern library (regex bucketing).** A series of regex patterns extracts salient address parts from highly variable 1945 strings. Matched rows are appended to the standardized set and removed from the active pool; unmatched residuals proceed to the catch-all.
6. **Catch-all tokenizer.** Remaining strings are split on commas/whitespace to create a coverage-complete token string.
7. **Abbreviation expansion layer.** For selected street-type or route tokens, both the short and long forms are appended (e.g., bwy, BROADWAY; ave, AVENUE; rt, ROUTE) to improve recall.
Policy update: prior to geocoding (Iteration 3), synonyms will be **deduplicated** with preference to the **long form** in the canonical text.
8. **Keywords & families.** The resulting text is saved as **keywords_1945** and split into NK/ASK/ALK families.
9. **Temporary canonical status.** For 1945 only, **keywords_1945** functions as the **temporary standardized address** until geocoding is applied in Iterations 2–3.

2.1.2 1975 workflow (light normalization + locality tokens)

- **Minimal cleaning.** Trimming, whitespace normalization, and ordinal spacing are applied.
- **Locality tokens.** The printed address is **augmented** with tokens parsed from the *Original Grantee City or Town* field to form **keywords_1975**. Quarter-section markers (e.g., NE/4) may therefore appear in the token stream at this stage; these will be revisited when the address-rulebook is formalized.
- **Token families.** **keywords_1975** is split into NK/ASK/ALK for matching.

2.1.3 1985 workflow (ZIP→city enrichment)

- **Minimal cleaning.** Trimming, whitespace normalization, and ordinal spacing are applied.
- **ZIP-based enrichment.** ZIP codes are mapped to **major city** names using a lookup; the city token is **injected** into **keywords_1985**. If a lookup fails, “Unknown” is inserted and a flag is retained—these rows remain in the dataset but are excluded from any **city-assisted** filters to avoid false positives.
- **Token families.** **keywords_1985** is split into NK/ASK/ALK for matching.

2.1.4 Case policy

Alphabetical tokens are normalized to **UPPERCASE** before token-family extraction to avoid case-driven mismatches; numeric tokens are unaffected. The original casing is preserved in the source column for audit and traceability. Later, Google/GeoPy formatted outputs will be stored exactly as returned by the APIs (mixed case), alongside the uppercase token keys used for joins.

2.1.5 1945 Pattern Library

The following table summarizes the major **address-pattern ideas** applied to the 1945 roll during Iteration 1. The entries describe *why* a rule exists, show **illustrative** raw snippets resembling those seen in 1945, and indicate the **transformed outcome** used to build **keywords_1945** and token families (NK/ASK/ALK). Examples are representative, not verbatim parcel records.

Pattern (purpose)	Raw 1945 example(s)	Iteration-1 transformation (keywords/tokens)
Leading house-number capture (anchors street records to a numeric NK token)	"123 BROADWAY", "1007 AVE Q"	Extract 123 / 1007 into NK; keep remaining tokens as ALK; produce keywords_1945 like "123 BROADWAY", "1007 AVE Q".
Ordinal spacing normalizer (handles compact ordinals)	"23RD ST", "4TH AVE"	Insert space → "23 RD ST", "4 TH AVE"; 23/4 become NK; RD/TH remain tokens for corridor logic.
City abbreviation expansion (locality recall)	"... LUB ..."	LUB → LUBBOCK (ALK); both forms may appear during expansion, later deduplicated to LUBBOCK before geocoding.
PO-box short form (box detection)	"BX 123", "BX-45"	BX → BOX; tokens retained as ALK/NK (123, 45).
Street-type short form: BROADWAY	"... bwy ..."	Append long form: "bwy, BROADWAY" in keywords_1945; later canonicalization keeps BROADWAY.
Street-type short form: AVENUE	"... ave ..."	Append long form: "ave, AVENUE"; later canonicalization keeps AVENUE.
Route short form: ROUTE	"rt 4", "on rt"	Append long form: "rt, ROUTE"; numeric (e.g., 4) → NK.
Building short form: BUILDING	"bld A"	Append long form: "bld, BUILDING"; A remains ALK.
Directional tokens (retain compass context)	"N 23 RD", "E AVE"	N, E, S, W, NE, SE, etc. retained as ASK (short alpha); assist corridor/side matching.
Punctuation-only placeholders (invalid rows)	"\", \"\\\"", ", "	Flagged/removed as invalid; excluded from keywords_1945.
Catch-all tokenizer (coverage for residuals)	"123, BROADWAY, LUB"	Split on commas/whitespace → tokens 123(NK), BROADWAY(ALK), LUB(→LUBBOCK); rejoin into keywords_1945.
Unknown locality from context (rare)	"..." with no clear city token	Tokens kept; locality unresolved until Iteration 2/3; row remains eligible for later geocoding.

Table 2.1 — 1945 Address Pattern Library and Transformation Rules

For comparison/matching, alphabetical tokens are normalized to **UPPERCASE** prior to NK/ASK/ALK extraction; original casing is preserved in provenance columns. Before geocoding (Iteration 3), synonym pairs in `keywords_1945` (e.g., "bwy, BROADWAY") are **deduplicated** with preference to the long form.

2.2 Iteration 2 — GeoPy-based normalization (archival years)

This iteration uses a geocoding service (via GeoPy/`geocoder.osm`) to transform cleaned address strings into a single **standardized formatted address** per row, using minimal but targeted preprocessing to maximize hit rates. At this stage the code captures the formatted address string; structured components and coordinates are not persisted. Audit/provenance columns from Iteration 1 remain intact.

2.2.1 1945 — preprocessing, geocoding, and retry logic

- **Inputs:** `Formatted_ADDRESS_1945` for non-missing rows; a unique `ID_1945` is generated for audit.
- **Pre-filters:** remove rows containing `BOX`, `RT`, `ROUTE`, `P0` (PO Box/route patterns not suitable for property-level geocodes).
- **Normalization:** ordinal join (e.g., `23 RD` → `23RD`) to match common OSM spellings.
- **Batching:** initial batches ~3,050 records; retries in smaller batches (~500).
- **Acceptance test:** accept only if the input's **first numeric token** (house number) is found within the returned formatted address; otherwise leave as missing for later passes.
- **Post-processing:** produce a file of geocoded results and a list of leftovers for further handling.

2.2.2 1975 — preprocessing, locality hints, and validation

- **Inputs:** `ADDRESS_1975`; if `Standardized_Address` existed from prior steps but lacked country/ZIP structure, it is re-evaluated.
- **Pre-filters for retries:** exclude rows containing `BOX`, `RT`, `ROUTE`, or `NaN` markers; drop cases where the last token cannot be a 5-digit ZIP.
- **Locality hints:** when geocoding, append **city context** by concatenating the address with a locality token (often derived from `ORIGINAL GRANTEE CITY OR TOWN` or explicitly ", Lubbock, TX, United States") to stabilize results.
- **Normalization:** ordinal join (e.g., `23 RD` → `23RD`).
- **Acceptance test:** enforce that the input's **leading house number** appears in the returned formatted address; otherwise reset to missing for future passes.

2.2.3 1985 — ZIP-anchored hints and staged retries

- **Inputs:** `ADDRESS_1985 + zip_code_1985`.
- **Pre-filters:** exclude `RT/route`, `P0/P.O./P 0`, and `box` rows; keep only plausible property-level strings.
- **Normalization:** ordinal join (e.g., `23 RD` → `23RD`).

- **Locality/ZIP hints:** if the address already contains "Lubbock", append ", TX, {ZIP}, United States"; otherwise append ", LUBBOCK, TX, {ZIP}, United States".
- **Dedup/cleanup:** remove accidental "TX, TX," duplicates.
- **Batching:** initial batches ~5,000; subsequent retries ~1,500.
- **Acceptance test:** accept only when the input's **leading house number** occurs in the returned formatted address; otherwise mark as missing and funnel to a retry path.
- **Targeted coercion for two-number patterns:** when an address consists of exactly two numeric tokens (e.g., "2301 23"), synthesize a coxing string by **inferring the corridor label from context** (e.g., tokens suggesting AVE/AVENUE/AVE Q/ROAD/HWY). The synthesized string takes the form "`{num1} {num2}{ordinal} {corridor}`, lubbock, tx {ZIP}" . If no corridor can be inferred confidently, do **not** force "street"; leave for later passes or geocoding with broader context. Track Success/Failure/NaN counters for this pass.

Example E-1985-1 — Two-number coercion

Input tokens: ADDRESS_1985 = "2301 23", zip code_1985 = 79410, context tokens include "AVE". Synthesized coaging string: "2301 23rd AVE, LUBBOCK, TX, 79410, United States". If the context had indicated "ROAD", the string would be "2301 23rd ROAD, ...". If no corridor token exists, the record is left for later passes rather than forcing "street".

2.2.4 Post-geocoding normalization(when present)

- Preserve the **verbatim GeoPy formatted address** in a dedicated audit column (e.g., `GeoPy_Address`).
- Create a **canonical text for matching** by stripping trailing ", United States" or county strings from the GeoPy result (e.g., `GeoPy_Canonical`).
- Subsequent iterations may pass `GeoPy_Canonical` forward; both audit and canonical forms are retained until Iteration 3 finalizes a single standardized field.

2.2.5 GeoPy Iteration outputs

Per year, this iteration appends/updates:

- **GeoPy_Address** (verbatim formatted address string from OSM; audit)
- **GeoPy_Canonical** (cleaned text for matching)
- Per-pass flags: acceptance based on house-number containment; retry counters where implemented
- For 1985: synthesized **Address_with_Zip** used in retries

Example E-GeoPy-1 — House-number containment acceptance

Input string	GeoPy_Address (returned)	Accepted?	Reason
"1007 AVE Q"	"1007 Avenue Q, Lubbock, Texas 79401, United States"	Yes	Leading house number 1007 present
"AVE Q NEAR 10TH"	"Avenue Q, Lubbock, Texas 79401, United	No	No leading house number to verify

	States"		
"2301 23"	"2301 23rd Avenue, Lubbock, Texas 79410, United States"	Yes	After coercion, 2301 appears; accepted

Table 2.2 — Example of GeoPy House-Number Containment Acceptance

2.2.6 Summary tables

Year	Input fields	Pre-filters before geocoding	Locality/ZIP hints appended	Batch sizes (approx.)	Timeout	Acceptance rule
1945	Formatted_ADDRESS_1945	Drop BOX/RT/ROUTE/PO ; non-digit leading tokens	None (property-level only)	3,050; retries 500	1.5s	Returned string must contain leading house number
1975	ADDRESS_1975 (+ locality tokens)	Drop BOX/RT/ROUTE; ensure ZIP-shaped last token when splitting	Append locality (often Lubbock, TX, United States)	5,000; retries 3,000	1.5s	Returned string must contain leading house number
1985	ADDRESS_1985, zip code_1985	Drop RT/route/PO/P.O./P O/box; non-digit leading tokens	Append TX + {ZIP}; insert LUBBOCK if absent	5,000; retries 1,500	1.5s–5s	Returned string must contain leading house number

Table 2.3 — GeoPy Preprocessing and Query Context (Archival Years)

Rule	Rationale	Applied in
Start-with-digit filter	Avoids PO Boxes, intersections, or non-situs strings	1945/1975/1985
Ordinal join (e.g., 23 RD→23RD)	Aligns with common OSM spellings	1945/1975/1985
Locality/ZIP hints	Stabilizes geocoding for ambiguous corridors	1975/1985
House-number containment check	Ensures the returned address still refers to the same property	1945/1975/1985
Strip trailing country/county (for canonical)	Removes extraneous tails from canonical text used in matching; retain verbatim in audit	As available
Coercion for two-number strings (context-aware corridor)	Converts ambiguous num num by inferring AVE/ST/ROAD/HWY from tokens; avoid forcing "street"	1985 (retry pass)

Table 2.4 — GeoPy Quality Assurance and Cleanup Rules

Stage	Audit columns retained	Canonical columns used for matching	Notes
Iteration 1 (regex/token)	Original address text; oldindex_year	keywords_year	1945 keywords_1945 acts as temporary canonical text.
Iteration 2 (GeoPy)	GeoPy_Address (verbatim)	GeoPy_Canonical (tails stripped)	GeoPy_Canonical is passed forward into Iteration 3.
Iteration 3 (Google)	Google_FormattedAddress (verbatim)	Final_Standardized_Address	Older canonicals may be pruned after finalization.

Table 2.5 — Audit and Canonical Column Lifecycle

2.3 Iteration 2.1 — Additional path: libpostal parsing + Nominatim standardization

The following shows an alternate/early Iteration-2 workflow that combines an **offline parser** (libpostal) with **online geocoding** (Nominatim via GeoPy). It reinforces the goals of: (i) creating a normalized string from irregular inputs, and (ii) obtaining a standardized formatted address suitable for downstream matching. This is to ensure high-quality parsing before geocoding, especially for 1945 irregular strings.

For 1945, the workflow loads the preprocessed file, drops earlier tokenization columns, and uses **Formatted_ADDRESS_1945** as input. Each string is first normalized with **libpostal** to regularize token order/spacing without external calls, then standardized via **Nominatim** (GeoPy), which returns a provider-formatted address or a sentinel on failure. This two-step path either confirms the parser output or backfills a canonical string when parsing is insufficient. A final cleanup trims country/county tails (e.g., “, United States”), preserving a verbatim audit string alongside a trimmed canonical for matching.

For 1975, the logic joins numeric ordinals (e.g., `23 RD → 23RD`) to match geocoder spellings, then constructs a geocodable line by appending a locality from *Original Grantee City or Town*, defaulting to “Lubbock” when missing. Because a missing street line can duplicate the locality (“Lubbock Lubbock”), a small guard is advisable to prevent duplication and insert a comma delimiter. The resulting standardized strings are then persisted for downstream steps.

Why two tools (libpostal + Nominatim)? The parser produces a **syntax-normalized** address string even when geocoding fails or times out, while Nominatim returns a **service-standardized** formatted address suitable as a canonical text. Using both increases overall coverage on noisy archival inputs.

2.4 Iteration 3 — Google Geocoding API canonicalization (archival years)

This iteration uses the Google Geocoding API to produce a **final standardized address** and to capture geospatial coordinates and granular components. The outputs form the canonical linkage key going forward.

For each submitted address, the code accepts the **first result** returned by `gmaps.geocode(address)` and records:

- **Google_Standard_address_{year}** — Google's `formatted_address` (mixed case).
- **latitude_{year}, longitude_{year}** — from `result[0]['geometry']['location']`.
Component columns — one column per `address_component type` in the result (e.g., `street_number_{year}, route_{year}, locality_{year}, administrative_area_level_2_{year}, postal_code_{year}, country_{year}`, and others present as needed).
- **County note:** the code **prints** whether `administrative_area_level_2 == 'Lubbock County'` but does **not** gate acceptance on this check; records outside the county are still written with a note in logs.

2.4.1 Acceptance & error handling

- **Acceptance policy:** if `gmaps.geocode(address)` returns at least one result, the first result is accepted and written; there is **no house-number containment test** or county gate at this stage.
- **Failures:** on exceptions, the function prints an error and returns a `Series([None])`, yielding missing component fields for that row.
- **Parallelism considerations:** multiple batches run concurrently; total QPS is shaped by the thread pool and the +5s pause per 50 calls within a worker loop.

2.5 Cross-Year Diagnostic — Matching

After Iterations 1–3 and the archival re-clean, the study verifies that key numeric attributes behave as true numerics and that canonical address strings are join-ready. The archival tables for 1945, 1975, and 1985 are each unique on `google_standard_address_{year}`; targeted fields are scanned for residual string contamination and profiled after coercion to float. Canonical addresses are then compared across years using the uppercase, tails-trimmed strings to summarize overlap sets (e.g., $1945 \cap 1975$, $1975 \cap 1985$, $1945 \cap 1985$). No rows are altered by these diagnostics. As an illustration, the address “1234 19TH ST, LUBBOCK, TX 79401, USA” is observed in 1975 and 1985 but not in 1945, a pattern reflected later in the archival presence flags.

Consideration	Applies to	Rule / Action	Example (Before → After)
Missing canonical	1945, 1975, 1985	Exclude rows with missing canonical key to fix the row	(blank / null) → row dropped

address		universe	
Numeric coercion (two-pass)	1945 value/tax; 1975 city/personal value & tax; 1985 assessed/tax	Convert numeric-looking strings → float; strip non-numeric decorations; non-numeric → NaN	"74,200" → 74200.0 · "A71000" → 71000.0 · "???" → NaN
Canonical comparison policy	1945, 1975, 1985	Compare by uppercase, tails-trimmed canonical strings	"1234 19th St , Lubbock, TX 79401 " → "1234 19TH ST, LUBBOCK, TX 79401"
Join-readiness punctuation	Archival → Contemporary join	Normalize _ to ", " when preparing cross-era joins	"1234 19TH ST_LUBBOCK_TX 79401_USA" → "1234 19TH ST, LUBBOCK, TX 79401, USA"
Overlap summary interpretation	1945, 1975, 1985	Summarize set membership across years (equality on canonical)	Example address: 1975 & 1985 only (absent 1945)

Table 2.6 — Data Cleaning and Join Readiness Rules for Archival Data

2.6 Handling repeated addresses within a year (multi-unit buildings)

Where multiple records share a canonical address within the same year, the records are aggregated by `google_standard_address_{year}` so that value and tax attributes form lists, while ZIP and coordinates retain the first non-null observation. A single coherent record per address is then selected by taking the maximum of a year-specific priority measure, with a tax measure as fallback when the priority list is entirely missing.

The rule is:

- 1945 prioritizes `value_of_city_property_1945` with fallback `total_tax_1945`;
- 1975 prioritizes `value_of_city_property_(total_county_value)_1975` with fallback `tax_total_(including_hospital_and_water_taxes)_1975`;
- 1985 prioritizes `assessed_1985` with fallback `gross_tax_1985`.

The chosen list index is applied across all parallel list-columns to maintain cross-field consistency, yielding scalar values throughout.

Outputs: one row per canonical address per year, plus repetition counts for audit. Ties, if encountered, inherit the order set during aggregation.

For example, a 1975 address with [68,500; 74,200; 72,900; 71,300] in `value_of_city_property_(total_county_value)_1975` resolves to the 74,200 entry; the same index is taken for owner and tax lists. Ties, if encountered, inherit the order established during aggregation. The resulting per-year outputs contain one row per canonical address together with repetition counts for audit.

2.7 Archival merge — 1945, 1975, 1985 (post-Iteration 3)

The de-duplicated year tables are consolidated into a single archival panel keyed by the canonical address. The union-distinct of `google_standard_address_{year}` defines the key set; each year's payload is attached by exact string equality via sequential left joins.

The merged table preserves year-suffixed fields, introduces boolean presence indicators for 1945, 1975, and 1985, and remains unique on the canonical key. Continuing the earlier example, an address present in 1975 and 1985 but absent in 1945 occupies one row with the 1975 and 1985 attributes populated and flags `present_1975 = TRUE`, `present_1985 = TRUE`, `present_1945 = FALSE`.

- Resulting structure: one row per canonical address; namespaced payloads by year; archival presence flags.
 - Deferrals: reconciliation of near-duplicate strings beyond canonicalization is reserved for later rulebook or spatial steps.
-

2.8 Contemporary years — standardization & geocoding (2012, 2020, 2021)

For each contemporary year, the workflow loads the standardized file and retains the core columns needed for linkage and valuation analysis: the printed situs address, total/land/improvement values, land size (acres/feet), and build year.

Address cleaning is deliberately light but strict enough to stabilize geocoding. The workflow drops obviously incomplete situs strings—cases that are only digits, a single word, or exactly two standalone words—and then appends a locality suffix to any remaining line that lacks it. If the string has neither “Lubbock” nor “TX,” it appends “, Lubbock, TX”; if it already includes “Lubbock” but not “TX,” it appends “, TX”; otherwise it leaves the string unchanged.

Examples.

- “1007 Avenue Q” → “1007 Avenue Q, Lubbock, TX”
- “3412 AVE D, Lubbock” → “3412 AVE D, Lubbock, TX”
- “5002 34th St, TX” → unchanged

Next, the workflow builds a time-series frame across the three contemporary years. It concatenates the post-clean addresses and deduplicates to create a unique address universe. Each year is then grouped by address, with the remaining fields aggregated into **lists** so repeated records at the same address are preserved as value arrays. An outer join across years on the address forms the inter-year panel; single-element lists are collapsed to scalars while true multi-value lists are retained.

Example.

For “3412 Avenue D, Lubbock, TX”: `TotalValue_2012` might be `[45,000, 70,000, 65,000,`

`55,000`, `TotalValue_2020 [118,000]`, `TotalValue_2021 [129,500, 131,200]`. (Downstream steps can select a consistent index when a single representative is required.)

Geocoding then runs over the combined frame using the Google Geocoding API. Addresses are processed in three sequential ranges (1–50k, 50,001–80k, and 80,001–end) in batches of ~3,000, with a 60-second pause between batches. For each address, the workflow records Google's formatted address (as the contemporary **standard address**) and the corresponding latitude/longitude, writing results back by row-index slice. There is no house-number containment test in this pass; any returned result is accepted.

Example.

“1007 Avenue Q, Lubbock, TX” → standard address “1007 Avenue Q, Lubbock, TX 79401, USA”, latitude ≈ 33.58..., longitude ≈ −101.84....

At this stage, the contemporary panel uses a single **standard address** column (no year suffix) as the canonical text for within-2012/2020/2021 joins. In contrast, the archival panel continues to use year-suffixed Google formatted address columns. A single unified canonical name (e.g., **Final_Standardized_Address**) is introduced only when the six-year dataset is assembled.

2.9 Cross-Era Matched Merge (1945/1975/1985 ↔ 2012/2020/2021) — Lean Checkpoint

To verify cross-era linkage, the study constructs a **conservative matched subset** by joining the archival and contemporary composites on **exact string equality** of standardized addresses. Archival punctuation is normalized from underscore separators to comma-space (,) to align with the contemporary **Standard_Address**; no fuzzy or spatial methods are used. The matched table retains the canonical address, a single coordinate pair from the contemporary side for consistency at this checkpoint, and year-suffixed payloads for 1945/1975/1985 and 2012/2020/2021.

A compact **Remarks** string lists the years present per row (e.g., “`1975 1985 2021`”). The output is unique on the canonical key, prints the row count with a small preview, and is saved as `merged_data.csv`. This exists purely to prove the mechanics before widening to the union panel.

2.10 Six-Year Union Panel — Consolidation & Harmonization

This step assembles a single address-keyed panel spanning **1945, 1975, 1985, 2012, 2020, 2021**. The approach takes the **union of unique addresses** across all sources, attaches year-specific payloads by exact string equality, and then harmonizes location fields (ZIP, latitude/longitude) into single columns suitable for mapping and downstream analysis.

2.10.1 Inputs and key set

The inputs are: (i) the one-row-per-address archival outputs from §2.7 (1945/1975/1985) and (ii) the contemporary composite from §2.8 (2012/2020/2021). Each source's canonical address is aliased to **Address**, concatenated, and reduced to a **union-distinct key set**.

- Union key size: **119,276** unique `Address` values.
- Intermediate snapshots are persisted after each stage to allow audit of shapes and column namespaces.

2.10.2 Sequential left joins (string equality)

The union key frame serves as a skeleton. Each source is then **left-joined** by `Address` without fuzzy tokens or spatial buffers. This preserves year-suffixed payloads and ensures that addresses present in only one or some years remain in the panel.

- Column namespaces remain unaltered (e.g., `value_of_city_property_1945`, `TotalValue_2021`).
- Merge order is recorded; the panel's row count remains fixed at the union size throughout.

Illustration. An address found in 1975 and 1985 but not elsewhere contributes one row; only the 1975 and 1985 fields are populated for that row at this stage.

2.10.3 ZIP extraction (two-pass refinement)

ZIP is derived from the canonical `Address` using a two-pass policy:

- **Pass A:** capture any 5-digit token anywhere in the string.
- **Pass B:** if `", USA"` is present, overwrite with the 5-digit token **immediately before `", USA"`** (if found).

Results: **119,216** rows have ZIP populated (of 119,276). Remaining cases lack an extractable 5-digit token in the required position.

Example. `"..., LUBBOCK, TX 79407, USA" → zip_code = 79407`; if no 5-digit token appears before `", USA"`, `zip_code` remains missing.

2.10.4 Coordinate unification & cleanup

Latitude/longitude are unified into single columns by preferring **archival coordinates** and filling from the contemporary frame only when archival values are missing. After unification, year-specific coordinate and ZIP columns are dropped.

- Fill order: `latitude_1945/longitude_1945` → `latitude_1975/longitude_1975` → `latitude_1985/longitude_1985` → contemporary `Latitude/Longitude`.
- Dropped after consolidation: all `latitude_{year}/longitude_{year}` / per-year ZIP columns and the contemporary `Latitude, Longitude` once merged.
- Final column order begins with: `Address, zip_code, latitude, longitude`, followed by the year-suffixed payload blocks.

2.10.5 Output structure and exports

The consolidated table preserves valuation/tax attributes by year and contemporary site attributes (lot sizes, YearBuilt) without recoding. The final structure is **119,276 rows × 66 columns**.

- Saved artifacts: a serialized checkpoint (pickle) and a CSV export of the consolidated panel.
- Grouped schema (for readability in code): location fields; 1945/1975/1985 value/tax blocks; 2012/2020/2021 valuation and site blocks.

2.10.6 Presence reading and a simple trajectory

Presence across years is inferred directly from non-missing fields. For clarity in narrative examples, the panel can be read as follows: if a row has populated 1975 totals and 1985 assessed values, but all 1945 fields are null, it is treated as present in 1975 and 1985 only. The same reading applies for contemporary years using `TotalValue_{year}`.

Trajectory example. Consider "1234 19TH ST, LUBBOCK, TX 79401, USA". If populated in 1975 and 1985 on the archival side and in 2021 on the contemporary side, the row's filled columns naturally trace a path **1975 → 1985 → 2021**.

3. Spatial Augmentation in ArcGIS Pro — Layers, Assumptions, and Prepared Classifications

This section documents the spatial work performed **after** consolidating the six-year panel in §2 (hereafter **LBK6Y Panel**). The objective is to enrich the panel with geography needed for targeted comparisons by: (i) digitizing a **possible redlining zone** identified by an external expert, (ii) aligning official **city limits** and **council districts**, and (iii) preparing point features and proximity rings for structured diagnostics.

Naming convention adopted. The consolidated table is referred to as **LBK6Y Panel**.

In ArcGIS Pro, the table is stored as `LBK6Y_tbl` and displayed as points as `LBK6Y_pts` using longitude/latitude.

3.1 External expertise and working assumption

A consultation was conducted with **Adam Pirtle**, staff attorney at **Legal Aid of NorthWest Texas (LANWT)** and a long-time Lubbock resident. After reviewing the methods and data sources, Adam indicated the **LBK6Y Panel** represented the most complete property dataset available for this inquiry.

Adam's public work focuses on zoning and neighborhood equity in Lubbock, including presentations and reporting that document the concentration of **industrial uses in North and East Lubbock** and their proximity to neighborhoods of color. For the purposes of this study, and **strictly as an operational assumption**, the highlighted area from his post is treated as a **possible redlining zone** to enable spatial contrasts in later sections. This is not a historical HOLC boundary and should be interpreted as a hypothesis informed by local expertise.

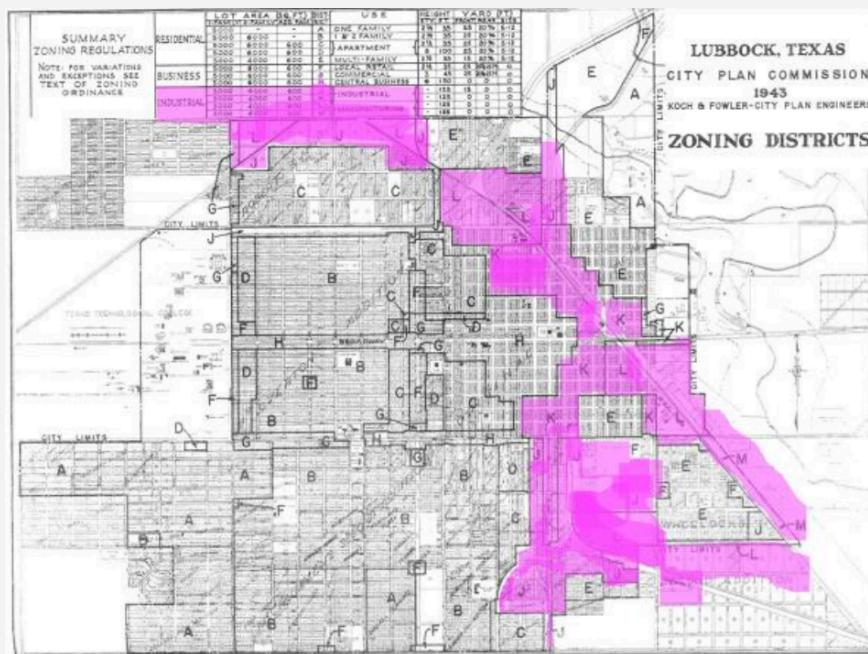


Figure 3.1 — Possible redlining zone (per Adam Pirtle, Legal Aid of NorthWest Texas)

3.2 Layers imported/prepared in ArcGIS Pro

The ArcGIS Pro project contains three primary layers plus derived buffers:

- **Possible redlining zone (polygon).** Feature class name: `RedliningZone`. Created by georeferencing Adam's shared figure and manually digitizing the highlighted area. Attributes include `source_org` ("Legal Aid of NorthWest Texas"), `source_author` ("Adam Pirtle"), `source_url` (public link), `source_capture_date`, `date_digitized`, and free-text `notes`.
- **City limits (polygon).** Feature class name: `CityLimit`. Downloaded from the City of Lubbock's official GIS/ArcGIS Online portal (maintained by City of Lubbock GIS & Data Services). Contains a `vintage_year` field set to 2024 (confirmed on download page); service metadata retained in project notes.
- **Council districts (polygons).** Feature class name: `CouncilDistricts`. Six non-overlapping districts that tile the city boundary when assembled.
- **LBK6Y points (points).** Feature class name: `LBK6Y_pts`. XY creation from `longitude/latitude` in the LBK6Y Panel table. Key fields: `Address`, `zip_code`, `latitude`, `longitude`, followed by year-suffixed payloads.

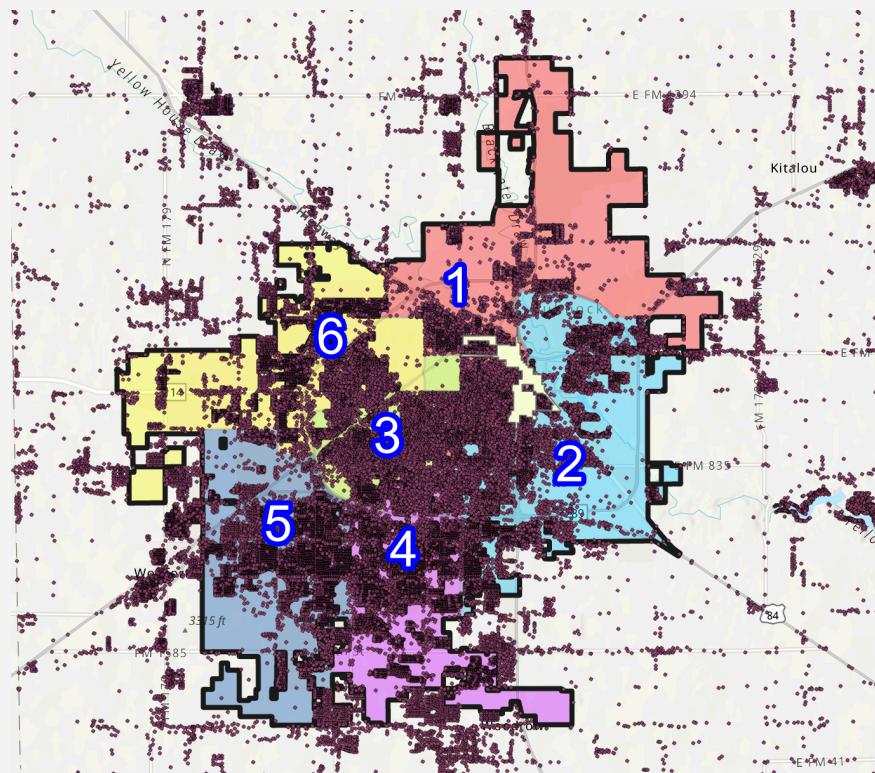


Figure 3.2 — LBK6Y points over CityLimit, Council Districts and Redlining Zone

3.3 Georeferencing and alignment

The redlining figure was georeferenced using a minimal set of **control points** placed on persistent intersections/landmarks visible in both the image and basemap. Residuals were checked to ensure the traced polygon aligned with the basemap to within visual tolerance. The city boundary and council districts were verified for **CRS compatibility** and reprojected as needed to a common working CRS.

- The digitized polygon was **dissolved** into a single feature where appropriate; interior holes were not introduced absent clear evidence in the source image.
 - The project metadata stores the control-point file and transformation notes for reproducibility.
-

3.4 Event layer and extents

`LBK6Y_pts` was plotted using the consolidated table's `longitude` and `latitude`. A visual sweep confirmed that most points lie **inside** the city boundary, with a non-trivial number **outside** the boundary. A quick count placed **inside-boundary points** at **~96,000** (precise count to be inserted after a scripted selection).

- **Count placeholder.** *Inside city limits: ≈ 96,000; Outside city limits: **[insert exact remainder]*.*
 - A formal selection by location will be run and logged; the counts will be updated in text once confirmed.
-

3.5 Proximity structure around the possible redlining zone

To study gradients around the hypothesized area, two **concentric buffers** were created around `RedliningZone`, measured from the **polygon boundary** (not centroid):

- **RedliningZone_Buffer_half_mile (0–0.5 mile).** Half-mile buffer outward from the redlining polygon's boundary. Points within the base polygon are referred to as the **core** for interpretation; points between the boundary and 0.5 miles fall in the half-mile ring.
- **RedliningZone_Buffer_one_mile (0.5–1.0 mile ring).** An additional half-mile ring **outside** the 0.5-mile buffer. This ring **excludes** the inner 0.5-mile zone.

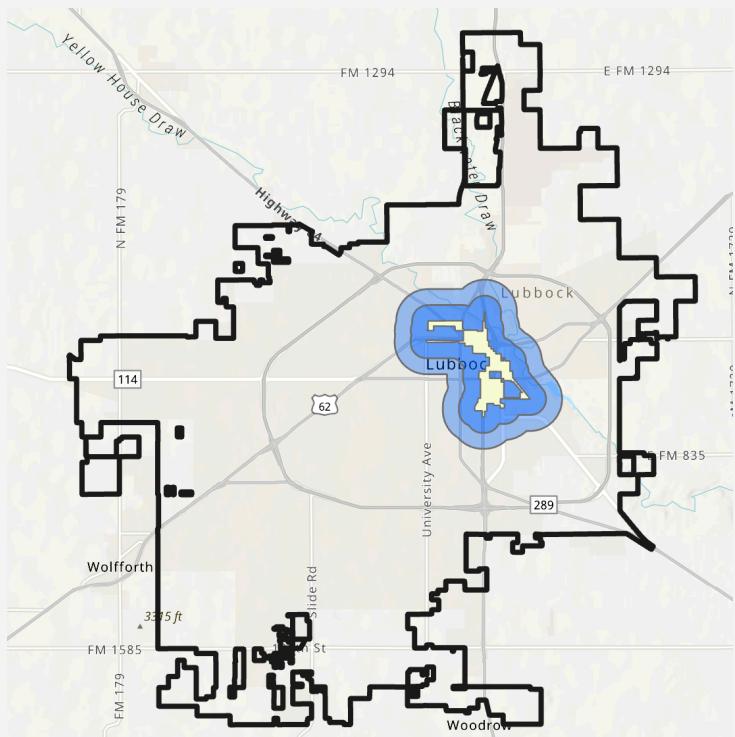


Figure 3.3 — Core polygon and two buffer rings (schematic over basemap)

3.6 Council districts

`CouncilDistricts` was added to support within-city stratification. Visual inspection confirmed that the six polygons form a **gap-free tessellation** that coincides with `CityLimit`. This layer will be used to annotate each property with a **district identifier (1–6)** for disaggregated summaries.

- Where boundary effects are material, later analyses can restrict to **inside-city** points or include a tolerance buffer to avoid misclassification near edges.
-

3.7 Prepared classifications added to LBK6Y points

The following fields were added to `LBK6Y_pts` via **Spatial Join / Select by Location** operations. Encoding follows a uniform convention of **1 = inside/true, 0 = outside/false**, except where noted.

- `IsInLubbock` — **1 if inside `CityLimit`, **0 if outside`.**
- `IsInCouncilDistrict` — integer **1–6** for points inside `CouncilDistricts`; **0** if outside all districts (i.e., outside city limits).
- `IsInRedliningZone` — **1 if inside `RedliningZone` (core polygon), 0 otherwise.**

- `IsInRedliningZone_half_mile` — **1 if within RedliningZone_Buffer_half_mile** (exclusive of the core), **0** otherwise.
- `IsInRedliningZone_one_mile` — **1 if within RedliningZone_Buffer_one_mile** (exclusive of the half-mile ring), **0** otherwise.

These attributes enable later summaries of valuations and site characteristics by proximity to the hypothesized zone and by district. Operation order and geoprocessing parameters are logged in project metadata (join type, match options, and selection predicates).

4. Analytical Findings (with Zoning Context)

Inputs for this section: the consolidated, cleaned six-year panel (1945, 1975, 1985, 2012, 2020, 2021) with standardized field names; five spatial/context flags; and zone proximity categories. Column renaming was completed upstream in **ArcGIS Pro** and is preserved throughout this analysis. All figures referenced below are provided as placeholders for insertion of final charts.

4.1 Study Frame and Inventory

The analytical frame comprises **119,184** property records, of which **96,592** ($\approx 81\%$) fall within Lubbock city limits. The leading ZIP codes by count are **79424 (21,055)**, **79423 (16,982)**, **79416 (11,160)**, **79403 (9,124)**, reflecting contemporary growth along the southwest and northwest corridors.

S.No	ZipCode	Count
1	79424	21055
2	79423	16982
3	79416	11160
4	79403	9124
5	79413	8137
6	79407	6891
7	79404	6045
8	79415	5502
9	79412	5070
10	79414	4994
11	79410	4203
12	79364	3712
13	79382	3692
14	79401	3536
15	79411	3250
16	79363	2270
17	79329	1359
18	79366	659
19	79311	621
20	79350	106

Table 4.1 — Top 20 ZIP Codes by Property Record Count

Data presence scales sharply by era: **1945 (2,186 rows)**, **1975 (26,346)**, **1985 (13,251)**, **2012 (99,599)**, **2020 (110,873)**, **2021 (113,054)**. The modern years dominate record coverage, while archival years provide targeted historical anchors.

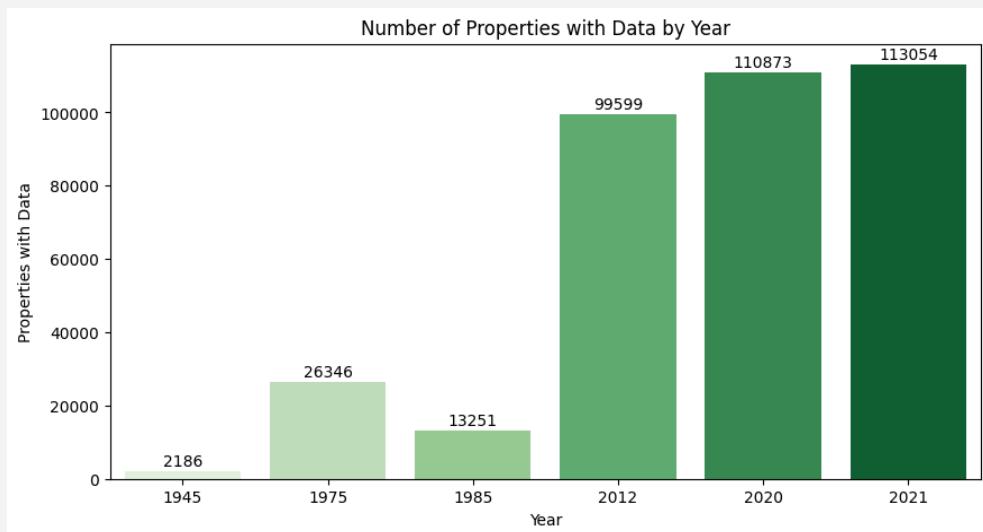


Figure 4.1 — Bar chart — properties with data by year

Council representation. Among contemporary years, district totals are balanced at scale: for 2021, Districts 1–6 register 14,517; 18,872; 13,226; 17,642; 15,820; 12,315 records respectively. Archival coverage is naturally thinner, but still most concentrated in Districts 1–3 for 1945–1985.

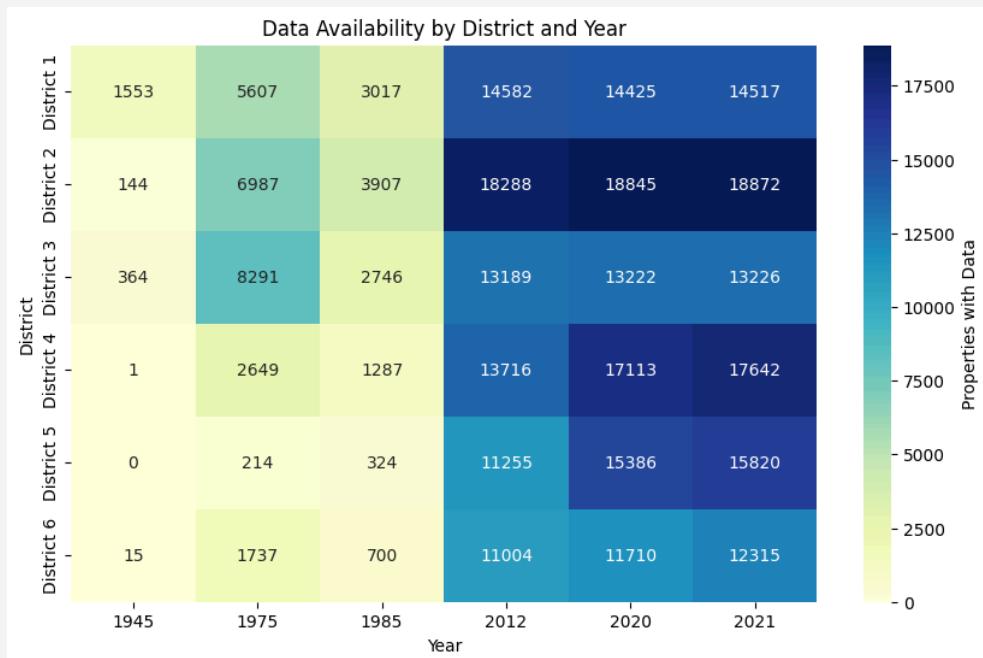


Figure 4.2 — Data availability by district and year

Redlining proximity. The enriched frame classifies parcels into **Redlining Zone**, **Half-mile Buffer**, **One-mile Buffer**, **Outside Both**. By 2021, counts are 876; 4,887; 5,349; 81,280 respectively. Redlined and buffer areas are concentrated almost entirely in Districts 1–2, with negligible or zero presence in Districts 3–6 in the archival years—consistent with historic city cores.

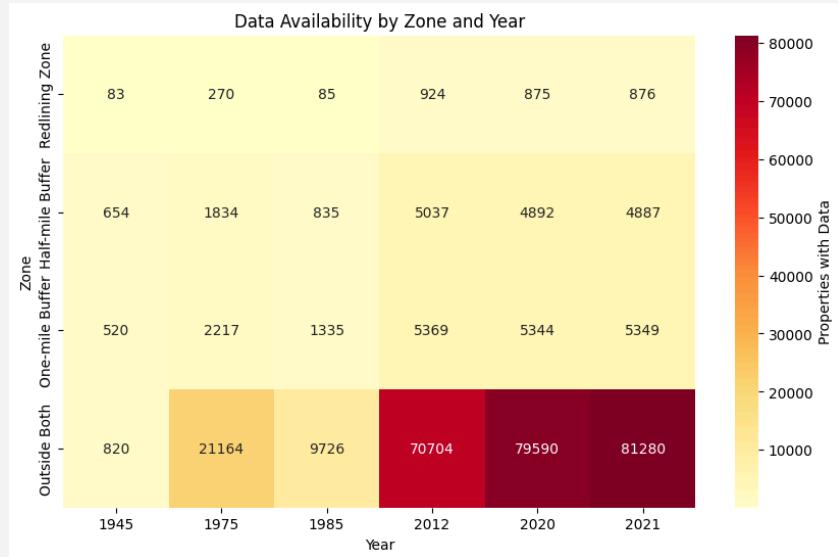


Figure 4.3 — Data availability by zone and year

4.2 Data Completeness and Readiness

Column-level completeness. Early-era fields (1945, 1975) exhibit notable sparsity (e.g., `DistrictSchool_1945` and `Area_1945` at ~100% missing), whereas contemporary core fields (2012–2021 values, land size, and most year-built fields) exceed 94% completeness. `YearBuilt_2021` is ~5.6% missing; parcel sizes in 2012–2021 are populated for >94% of in-city records.

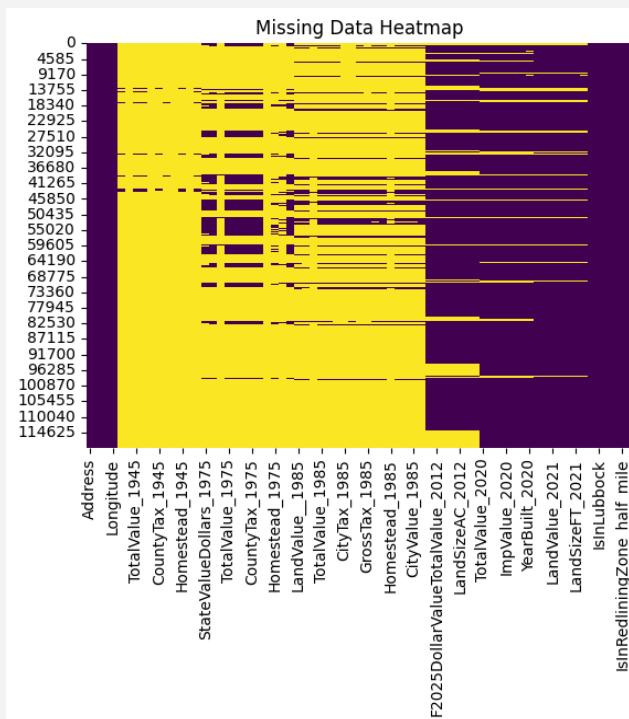


Figure 4.4 — Heatmap of missingness by column and year

4.3 Spatial Coverage — Districts × Zones

District 1 & 2 are the locus of historical zoning effects.

- In 2021, District 1 contains ~596 parcels in historically redlined areas, with 3,636 in the half-mile buffer and 3,390 in the one-mile buffer. District 2 shows 280 redlined, 1,248 half-mile, 1,957 one-mile.
- Districts 3–6 have no parcels flagged inside the defined redlining footprints, underscoring their later development vintage and the east–west spatial skew of historical policy lines.

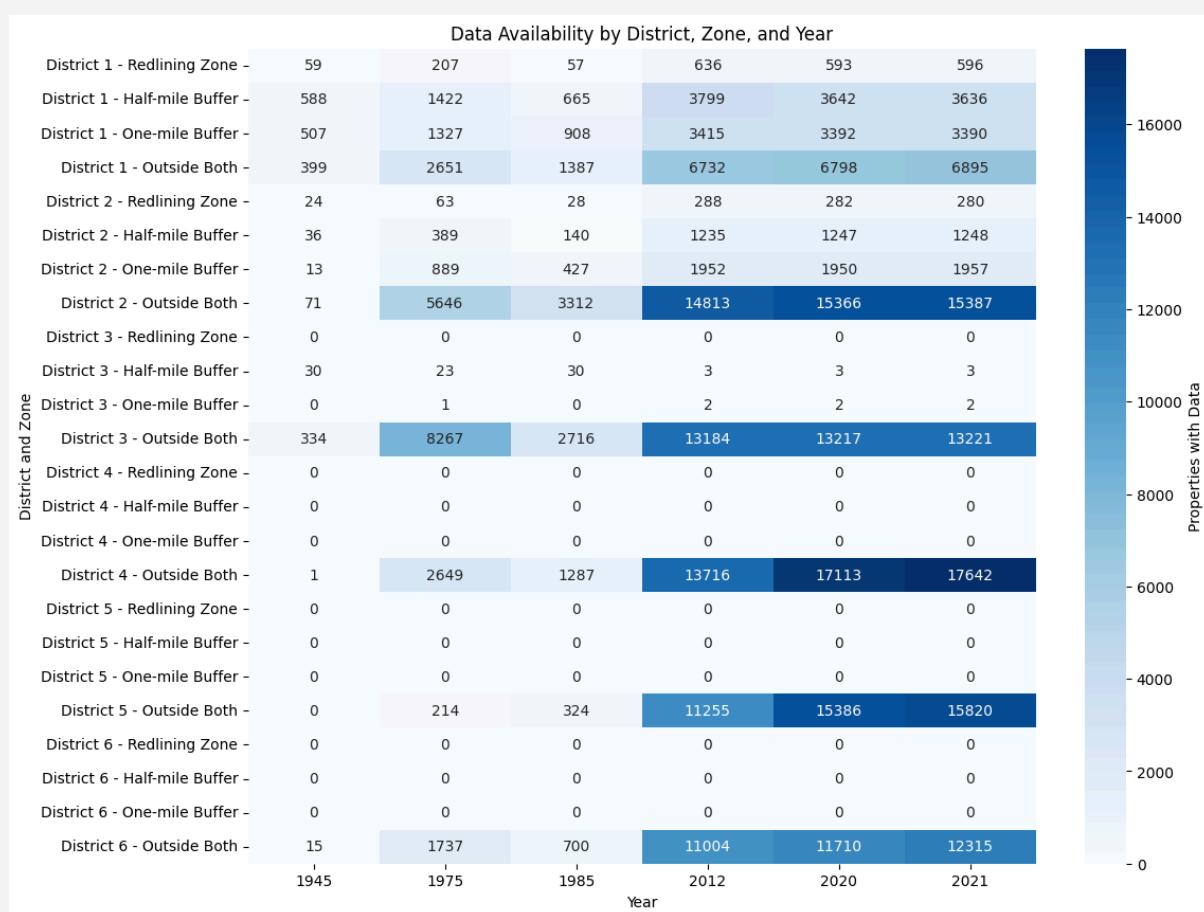


Figure 4.5 — Small multiples — district×zone counts by year

The zone distribution aligns with well-documented urban patterns: older center-city districts concentrate the historical designations; subsequent suburbanization expands inventory in “Outside Both,” which dominates by 2012–2021.

4.4 Value Structure and Cross-Year Correlations (diagnostic)

The cross-metric correlation structure validates internal consistency by era and clarifies comparability:

- **Within-era coherence (1975 & 1985).** Value measures and composite tax totals move almost one-for-one, reflecting a common appraisal base. For 1985, [CityValue_1985](#), [CountyValue_1985](#), [SchoolValue_1985](#), and [GrossTax_1985](#) are **highly collinear**, as expected under coordinated assessment regimes.
- **Cross-era drift.** Correlations between archival values (1945/1975/1985) and contemporary totals (2012/2020/2021) are **modest**—partly due to address churn, redevelopment, boundary changes, and method shifts. This argues for **within-era comparisons** and **growth rates** rather than raw cross-era level comparisons.
- **Year built.** [YearBuilt_{2012–2021}](#) exhibits **weakly negative** association with the presence of very old high values (selection and survivorship) and **positive** association with contemporary improvements—hallmarks of infill versus greenfield dynamics.

4.5 Property Value Levels by District (outlier-trimmed)

Using the outlier-trimmed summaries:

- **2012–2021 levels.** District 4 posts the highest adjusted averages (e.g., **\$209k in 2012 → \$251k in 2021**), followed by **District 5 (\$197k → \$238k)** and **District 3 (\$123k → \$153k)**. Districts 1–2 remain lower in level (**\$62–86k** ranges in 2012–2021), consistent with older stock and smaller lot patterns.

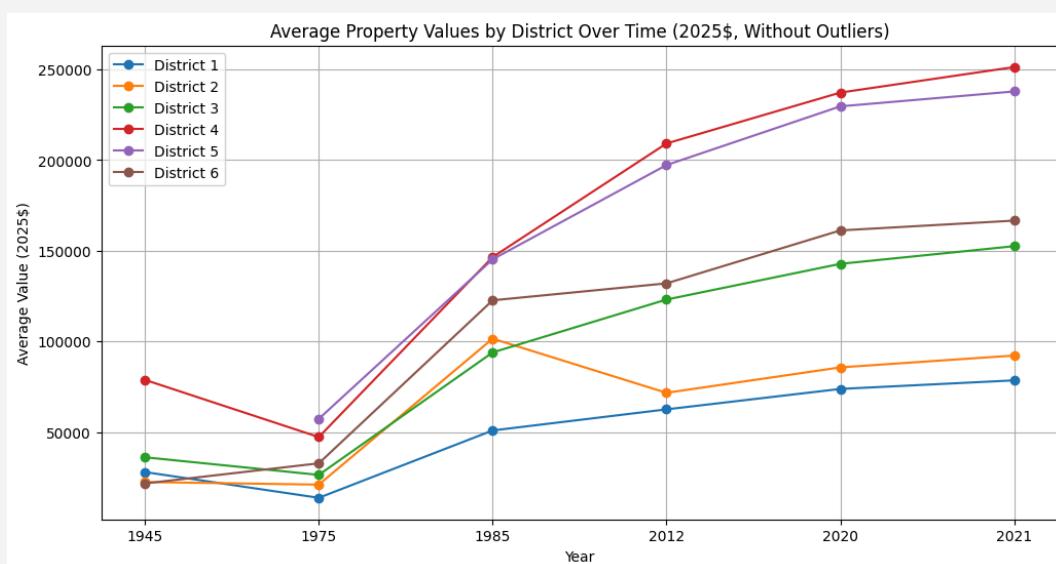


Figure 4.6 — Line chart of district averages by year

- **Archival anchors.** In 1945, measurable records clustered in Districts 1 & 3 (means ≈\$28–36k in inflation-adjusted dollars), with District 4 represented by a single high record (≈\$78.9k). By 1985, District 2–6 show sharp divergence, reflecting citywide expansion and

valuation spread.

The level gap between Districts 1–2 and 4–5 persists into the contemporary period even after trimming outliers, suggesting structural differences in housing stock, parcel sizes, and location amenities rather than mere data noise.

4.6 Property Value Levels by Redlining Proximity (outlier-trimmed)

- **Outside Both** dominates levels and counts. In 2012, adjusted averages are ≈\$135k (Outside) versus \$51k–65k inside the designated zone or buffers. By 2021, the pattern persists: **Outside Both** ≈ \$173k, **Redlining Zone** ≈ \$68.9k, **Half-mile** ≈ \$50.1k, **One-mile** ≈ \$65.4k.
- **Archival eras mirror the geography.** In 1975 and 1985, “Outside Both” far exceeds the zone/buffer averages, reflecting the newer, larger parcels built beyond the historical core.

Proximity to historic redlining designations remains associated with **lower contemporary value levels**, net of extreme outliers. This is consistent with legacy underinvestment and smaller lots building up around the original core.

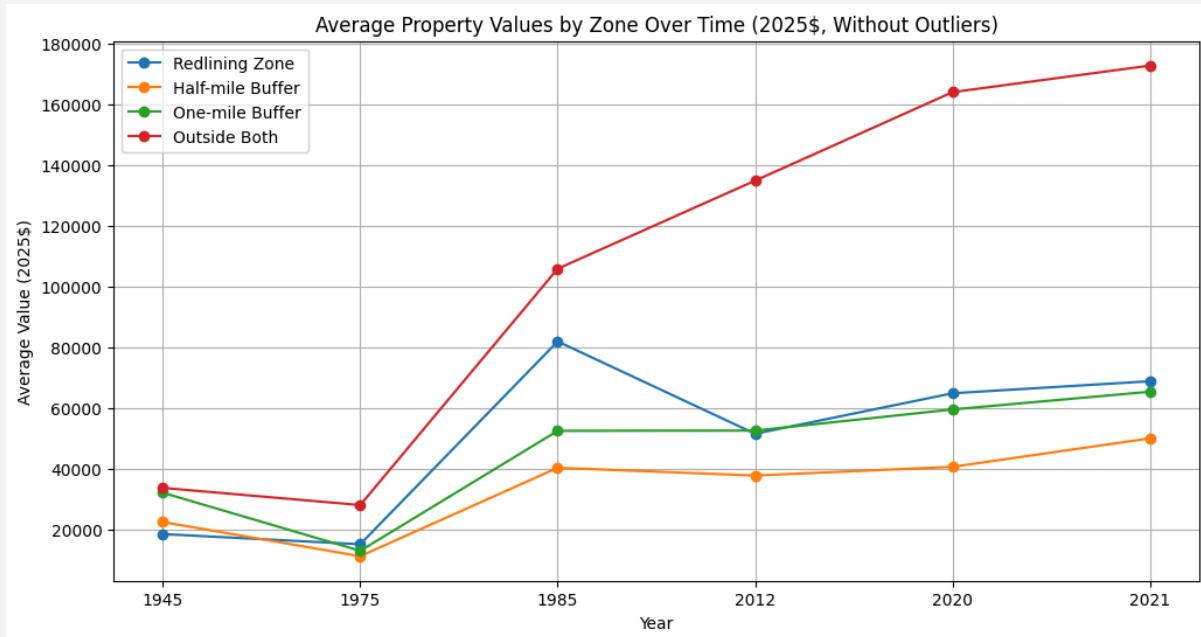


Figure 4.7 — Grouped lines by zone

4.7 Growth Dynamics by District (percentage change in adjusted means)

Growth rates (outlier-trimmed) illuminate the timing of value change:

- **1945 → 1975:** Districts 1–4 show declines in adjusted averages (e.g., D1 **-50.5%**, D4 **-39.9%**), while D6 increases (**+50.8%**). The declines likely reflect **coverage drift** (which addresses are present each year), **method changes** in valuation between tax eras, and mid-century neighborhood aging.
- **1975 → 1985:** All measured districts surge (**+209% to +380%**), consistent with **1980s appraisal inflation and city expansion**.
- **2012 → 2020 → 2021:** Broad, moderate appreciation across districts (**~13–22%** to 2020; additional **3–7%** into 2021). Districts 4–5 appreciate from a higher base; Districts 1–2 maintain lower bases but still register steady gains.

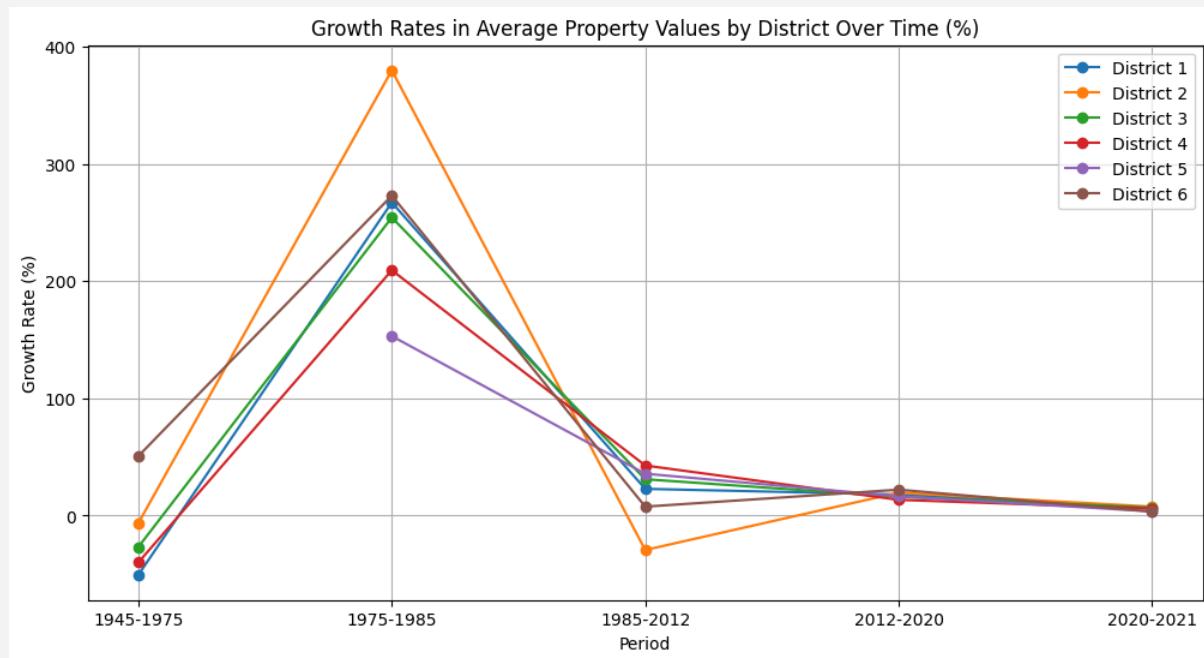


Figure 4.8 — Waterfall panels for district growth intervals

Caution. Early-era percent changes are sensitive to **sample composition** (which addresses persist across eras) and should be read as **indicative** rather than exact cohort appreciation.

4.8 Parcel Price per Square Foot (PPFS)

Coverage. Non-null PPFS counts: **1975 (401)**, **1985 (11,896)**, **2012 (80,341)**, **2020 (89,130)**, **2021 (90,763)**. Missingness stems from parcels lacking size or value for the given year.

Distributional shape. Medians track a plausible trend—**~\$15.09 (2012) → \$18.70 (2020) → \$19.99 (2021)**—while means sit higher (**\$16.8 → \$26.1 → \$27.8**) due to **right-tail outliers**. Reported maxima (e.g., **\$419,398** in 2020; **\$448,181** in 2021) are mechanical artifacts of **near-zero parcel areas or data entry quirks** and are excluded from level comparisons.

By zone (2021). PPFS medians are lowest in the **Redlining Zone (~\$5.34)** and **Half-mile Buffer (~\$7.15)**, intermediate in the **One-mile Buffer (~\$8.71)**, and highest **Outside Both (~\$21.72)**. Means are more inflated in buffer areas (e.g., Half-mile mean $\approx \$36.16$) due to a small number of extreme micro-parcels.

4.9 Quality Notes and Analytical Guardrails

1. **Renamed columns (ArcGIS Pro).** Standardized, intuitive field names were finalized *prior* to analysis. A small number (<5) relied on informed judgment; these are flagged in the data dictionary and do not affect numerical content.
 2. **Era comparability.** Differences in tax rules, appraisal bases, and spatial coverage across 1945–1985 versus 2012–2021 limit strict level comparisons. Prefer **within-era** structure and **interval growth** over raw cross-era levels.
 3. **Zone definitions.** Zone membership (**Redlining Zone**, **Half-mile**, **One-mile**) is **binary at the parcel centroid**; parcels near boundaries may experience measurement jitter across buffering thresholds.
 4. **Outliers.** All level statements use the provided **outlier-trimmed** summaries where applicable; for PPFS, extreme tails arise when land area approaches zero.
 5. **Counts vs. coverage.** Graphs should annotate **N by group** (district×zone×year) to contextualize comparisons—especially in 1945 and 1975 where counts are small.
-

4.10 Executive Synthesis

- The **location of historical redlining** continues to **predict lower contemporary values** and **lower PPFS medians**, even after trimming extreme outliers.
- **Districts 4–5** lead the city on adjusted value levels and maintain steady growth into 2021; **Districts 1–2** appreciate more slowly from lower bases, mirroring the housing stock and lot size profiles in the east-central core.
- Cross-era comparisons are most reliable when framed as **interval growth within districts** and **relative differences by zone**, not as raw level equivalences across appraisal regimes.

5. Per-Square-Foot (PPFS) Metrics — Construction, Imputation, and Findings

This section formalizes how Parcel Price-Per-Square-Foot (PPFS) is built across six study years, documents imputation choices for missing lot sizes, applies inflation adjustments to constant 2025 dollars, and interprets distributional patterns by redlining proximity and district. It is written to be auditable and to stand alone for reviewers who focus on §4 results but want to see the numeric provenance behind PPFS.

5.1 Data inputs and PPFS definition

Inputs. Consolidated six-year panel with zone flags and standardized names (119,184 rows). Core fields used here:

- **Total values:** `TotalValue_{year}` (1945, 1975, 1985, 2012, 2020, 2021) and their inflation-to-2025 counterparts `F2025DollarValueTotalValue_{year}`.
- **Lot size:** `LandSizeSqFT_{year}` for 2012/2020/2021 (directly observed). For 1945/1975/1985 the lot areas are **not** reliable in the source (see §5.2), so we impute selectively from the contemporary years (see §5.3–§5.4).
- **Zone flags:** `IsInRedliningZone`, `IsInRedliningZone_half_mile`, `IsInRedliningZone_one_mile`.

Definition. For a given year t ,

$$\text{PPFS}_t = \text{TotalValue}_t \div \text{LandSizeSqFT}_t$$

(units: \$/sqft; we present both nominal and 2025-dollar versions)

Where the value is inflation-adjusted, we denote `F2025TotalValuePerSqFT_{t}`.

5.2 Source caveat — archival “area” is not a usable measurement

External source inspection (microfilm printouts) and internal checks show the 1985 “Area” fields behave as **administrative codes** repeated across thousands of parcels rather than physical sizes; 1945 area is entirely missing; 1975 area is ≈98% missing. Consequently, archival area columns are dropped from PPFS construction. (§4 noted the same limitation when discussing completeness.)

5.3 Indicator rows for year-membership

We tag rows that actually **belong** to a year with `BelongTo_{year}` (1 if that row has the year’s total value, else 0). Counts:

- `BelongTo_1945` = **2,186** rows
- `BelongTo_1975` = **26,346** rows
- `BelongTo_1985` = **13,251** rows
- `BelongTo_2012` = **99,599** rows
- `BelongTo_2020` = **110,873** rows
- `BelongTo_2021` = **113,054** rows

These flags gate where we attempt lot-size backfills for archival years (we only backfill when a parcel is present in that year's value universe).

5.4 Lot-size imputation policy (archival years only)

Logic. For parcels that belong to 1945/1975/1985, we copy the **best available** contemporary lot size in the following order: `LandSizeSqFT_2012` → `LandSizeSqFT_2020` → `LandSizeSqFT_2021`. We also record the **source year** in `LandSizeSqFT_ImputationSource`.

Source distribution across the full panel.

- From **2012**: **97,072** parcels
- From **2020**: **13,456** parcels
- From **2021**: **2,497** parcels
- **Still missing after all sources**: **6,159** parcels

Selective results (imputed only for rows belonging to the year):

- **1945**: filled lot size for **1,578** of 2,186 rows (608 still missing)
- **1975**: filled lot size for **23,553** of 26,346 rows (2,793 missing)
- **1985**: filled lot size for **12,251** of 13,251 rows (1,000 missing)

Lot sizes generally remain stable at the parcel level absent subdivision or assemblage. Using the most recent reliable area as a proxy for earlier years is a standard, conservative approach for historic value-per-area measures, especially when the alternative is unusable archival “area” codes.

5.5 PPFS availability and inflation to constant dollars

Once lot sizes are available, we compute PPFS by year. Non-null PPFS counts (post-imputation):

- 1945: **1,578**
- 1975: **23,553**
- 1985: **12,251**
- 2012: **97,072**
- 2020: **108,634**
- 2021: **110,932**

Inflation factors to 2025 dollars (applied multiplicatively to PPFS):

- 1945 → ×**17.53**

- 1975 → **×5.87**
- 1985 → **×2.93**
- 2012 → **×1.37**
- 2020 → **×1.22**
- 2021 → **×1.16**

All summary statements below are available both in nominal \$ and 2025\$; visualizations prefer 2025\$ when mixed years are shown.

5.6 Distributional shape and citywide drift (2012–2021)

Citywide PPFS exhibits a right-skew with a thin extreme tail driven by sliver parcels or data entry at very small areas. Medians move in a tight band while means sit higher:

- **Medians:** ~\$15.09 (2012) → \$18.70 (2020) → \$19.99 (2021)**
- **Means:** ~\$16.8 → \$26.1 → \$27.8

Interpretation: the median series reflects typical residential parcels; the widening mean-median gap is consistent with (i) a small number of near-zero-area entries, (ii) downtown/commercial parcels with high improvement values on modest footprints, and (iii) normal dispersion during appreciation cycles. In narrative and figures we emphasize **medians and IQRs**; extreme outliers are retained but visually down-weighted (log scale) and flagged in notes.

5.7 Spatial contrasts — PPFS by redlining proximity (2021)

Using mutually exclusive zone categories (Redlining Zone, Half-mile Buffer, One-mile Buffer, Outside Both), the **2021** PPFS medians follow a clear gradient:

- **Redlining Zone:** ~\$5.34/sqft
- **Half-mile Buffer:** ~\$7.15/sqft
- **One-mile Buffer:** ~\$8.71/sqft
- **Outside Both:** ~\$21.72/sqft

The Outside-Both category dominates counts and levels, reflecting newer stock and larger-scale subdivision beyond the historic core. Buffer areas show higher **means** than their medians (e.g., Half-mile mean ≈ **\$36.16**), indicating a handful of small-footprint/high-value parcels; therefore, **median/IQR** is the preferred central tendency for zone comparisons.

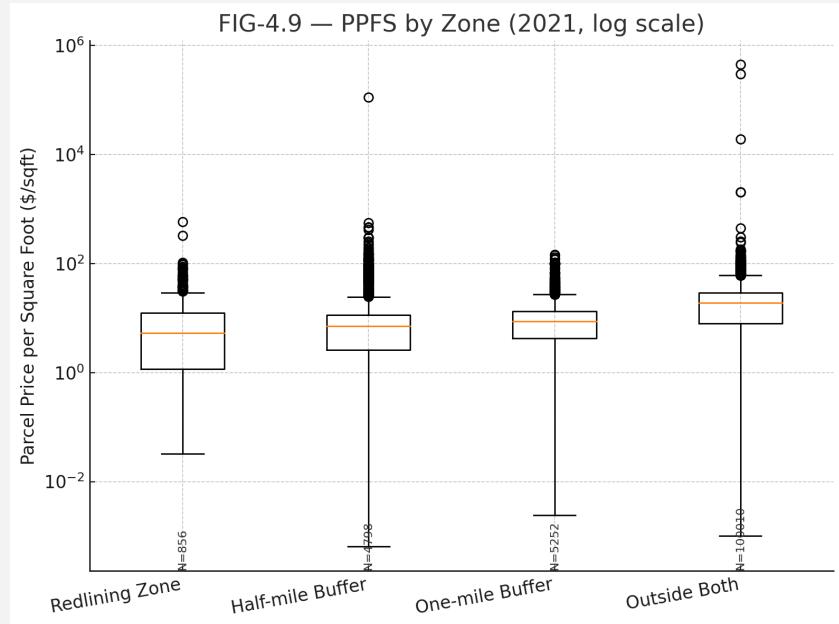


Figure 5.1 — PPFS box-and-whisker by zone 2021 (log y-axis)

The log scale stabilizes the right tail while preserving rank differences among medians. Annotation of **N** per box is essential.

Why the dip near the core? Lower medians inside the redlining footprint and its buffers align with:

1. older/denser housing stock with smaller improvements relative to land;
2. deferred investment patterns and amenity gaps that persisted through later decades;
3. smaller typical lot sizes (denominator effects); and
4. selection effects—properties that remain in the archival core differ systematically from post-2000 subdivisions outside both buffers. Where a zone's line sits **below** others in any table or figure, it is usually explained by a mix of these factors rather than one single driver.

5.8 District cross-sections (brief)

Consistent with §4, Districts **4–5** lead on levels into 2021, with Districts **1–2** lower in level but registering steady appreciation. Within any district, the zone gradient typically holds—parcels outside both buffers post higher PPFS medians than those inside/buffered.

5.9 Quality guardrails and edge cases

- **Zero/near-zero areas.** Observations with area ≤ 0 are excluded from PPFS; very small areas are retained but naturally appear as high-value outliers. Use log scaling and median/IQR summaries to limit influence in interpretation.
- **Imputation scope.** Backfills for 1945/1975/1985 operate **only** where the row belongs to that year (presence of that year's value). This avoids inventing PPFS for parcels that have no valuation context in that year.

- **Audit trail.** The column `LandSizeSqFT_ImputationSource` records the source year for every copied lot size. Counts: 2012 (97,072), 2020 (13,456), 2021 (2,497), missing (6,159).
 - **Comparability.** PPFS mixes land and improvements by design (a tax-roll total divided by land area). For asset-mix analysis, pair PPFS with `ImpValue_{year}` shares or run sensitivity using improvement values only.
-

5.10 Takeaways

1. The PPFS construction is transparent and reproducible, with archival area codes explicitly excluded and a documented fallback ladder for lot sizes.
2. Citywide medians rose moderately 2012–2021 while the right tail thickened—an expected pattern when small-footprint parcels appreciate faster.
3. Proximity to historical redlining remains associated with **lower** PPFS medians in 2021, even after outlier control and on a log scale.
4. For policy audiences, report **median** levels by zone and district, and display **FIG-4.9** with N-labels. Outlier maxima should be footnoted, not emphasized.

6. Main PPFS Analysis (Nominal and Inflation-Adjusted to 2025\$)

This section delivers the core **parcel-price-per-square-foot (PPFS)** analysis across six benchmark years (1945, 1975, 1985, 2012, 2020, 2021). We report nominal \$ and constant 2025\$ values, contrast **All parcels** with the **Complete** cohort (N=281 parcels with valid PPFS in *all* six years), and structure results by zone proximity and council district. All numbers below are **directly computed** from the imputed master CSV (see §5 for PPFS construction and inflation to 2025\$).

Inputs. `csv` with imputed lot sizes for archival years (§5) and pre-computed PPFS columns: `TotalValuePerSqFT_{year}` (nominal) and `F2025TotalValuePerSqFT_{year}` (inflation-adjusted). We derive a mutually exclusive **ZoneCategory** with priority: **Redlining Zone** → **Half-mile Buffer** → **Between 0.5–1 mile** → **Outside Both**. District labels are taken from the `IsInCouncilDistrict` field where present.

Unless otherwise stated, y-axes are PPFS in **2025 dollars**; group comparisons annotate **N** and, where tails are long, a **log-y** companion is provided. Download links for every figure and table are included inline.

6.1 Cohorts and Coverage

Panel size (inside Lubbock where flagged):

- **All parcels** with non-null PPFS-2025\$ by year: **1945 = 1,542; 1975 = 23,277; 1985 = 11,436; 2012 = 80,341; 2020 = 89,130; 2021 = 90,763** (TAB-6.1).
- **Complete cohort:** **281 parcels** with valid PPFS-2025\$ in *all* six benchmark years.

Zone distribution. Outside-Both dominates coverage; core and buffer rings concentrate nearer the historical center. Composition tables (TAB-6.2) provide Ns and shares by cohort.

Year	Cohort	N
1945	281-Complete	281
1945	All	1542
1975	281-Complete	281
1975	All	23277
1985	281-Complete	281
1985	All	11436
2012	281-Complete	281
2012	All	80341
2020	281-Complete	281
2020	All	89130
2021	281-Complete	281
2021	All	90763

Table 6.1 — Property Cohort Counts by Year

Cohort	Zone	N	Share
All	Outside	83639	1
All	Half-mile Buffer	5959	0
All	Between 0.5-1 Mile	5843	0
All	Redlining Zone	1151	0.01191610071
281-Complete	Outside	141	0.501779
281-Complete	Between 0.5-1 Mile	87	0.3096085409
281-Complete	Half-mile Buffer	52	0.1850533808
281-Complete	Redlining Zone	1	0

Table 6.2 — Zone Composition by Property Cohort

6.2 Definition and Units

- **PPFS (nominal):** $\text{TotalValue}_t \div \text{LandSizeSqFT}_t$ (in \$/sqft at year t).
 - **PPFS (2025\$):** nominal PPFS multiplied by the CPI-based inflation factor from \$5 to restate values in 2025 dollars.
 - **Guardrails:** Exclude non-positive areas; emphasize medians and IQR due to right-skew; provide log-y views and CCDFs for tail inspection.
-

6.3 Citywide Levels (2025\$) — All vs Complete

Medians (PPFS-2025\$):

- **All parcels:** 1945 = \$4.19, 1975 = \$2.78, 1985 = \$11.01, 2012 = \$15.09, 2020 = \$18.70, 2021 = \$19.99 per sqft.
- **Complete (281):** 1945 = \$4.40, 1975 = \$2.10, 1985 = \$7.52, 2012 = \$13.28, 2020 = \$14.80, 2021 = \$15.74 per sqft.
(See TAB-6.3 for IQRs, p90s, mins/maxes.)

The citywide (All) distribution is more spread with heavier upper tails than the 281 cohort. The level gap (All > 281) is visible in most years and widens in more recent cross-sections, reinforcing the value of reporting medians and IQR rather than means.

Year	Cohort	N	Median	Q1	Q3	P90	Mean	Min	Max
1945	281-Complete	281	4	4	5	6.26	5	0.082494	179.5072
1945	All	1542	4	3.30	5	6.40	97	0.020869	142431.3
1975	281-Complete	281	2	1.78	2.62	3.42	2	0.092441	14.44923
1975	All	23277	2.776112	1.819625	4.275394	5.464724	7.952626	0.000449	108228.1

1985	281-Complete	281	7.523403	6.333396	9.093464	11.35438	9.505267	0.110306	131.6207
1985	All	11436	11.00569	6.396296	16.30409	20.75881	12.50828	0.00077	756.4484
2012	281-Complete	281	13.2785	9.453196	17.69504	22.27884	15.07421	0.174623	113.1546
2012	All	80341	15	7.73	22.95	31.52	17	0	482.6155
2020	281-Complete	281	15	10.69	22	30.21	17	0.155513	129.4551
2020	All	89130	19	9.25	28.50	38.10	26	0	419397.9
2021	281-Complete	281	15.74219	11.64786	22.37226	30.047	18.76533	0.147865	134.004
2021	All	90763	19.99404	10.0775	30.11357	39.95333	27.78401	0	448180.5

Table 6.3 — PPFS (2025\$) Summary Statistics by Year and Cohort

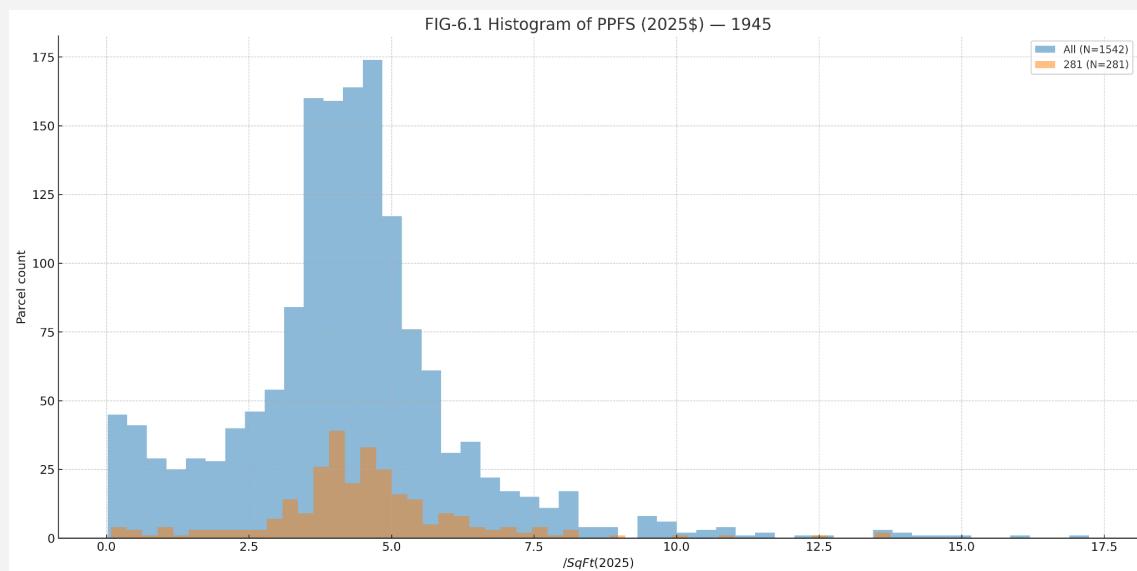


Figure 6.1 — Histograms (2025\$; All vs 281) — 1945

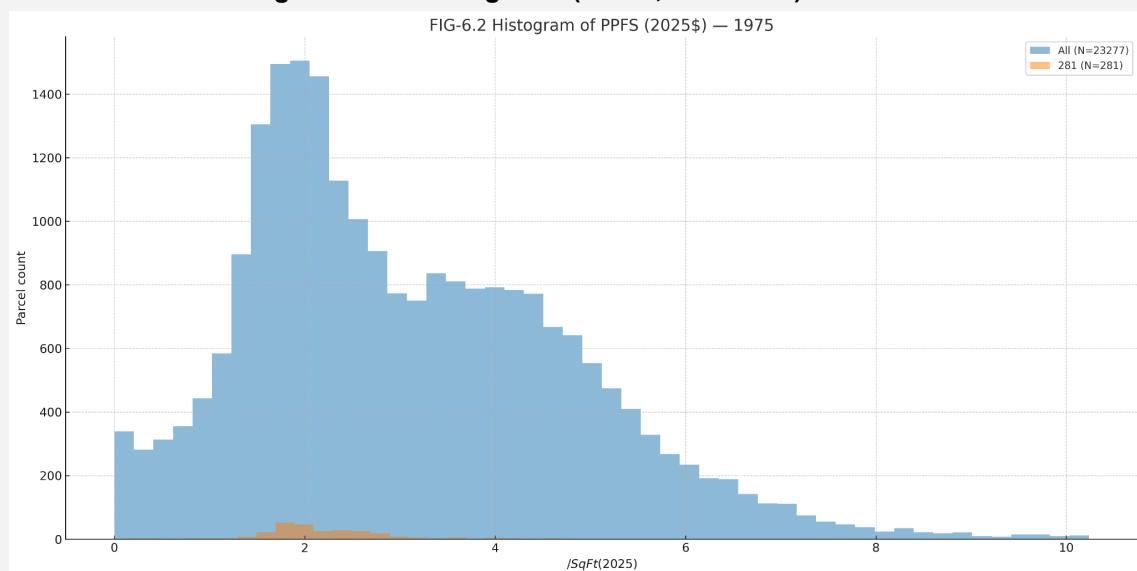


Figure 6.2 — Histograms (2025\$; All vs 281) — 1975

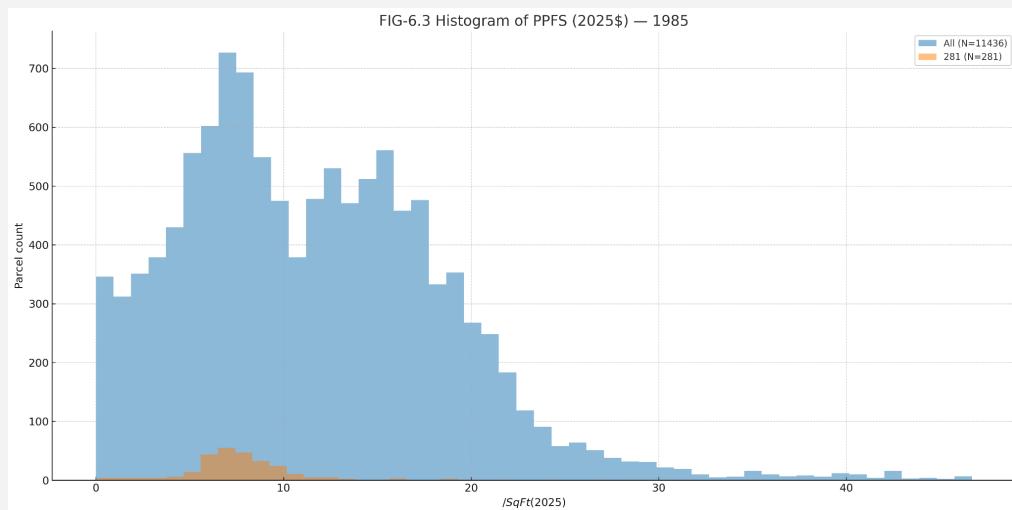


Figure 6.3 — Histograms (2025\$; All vs 281) — 1985

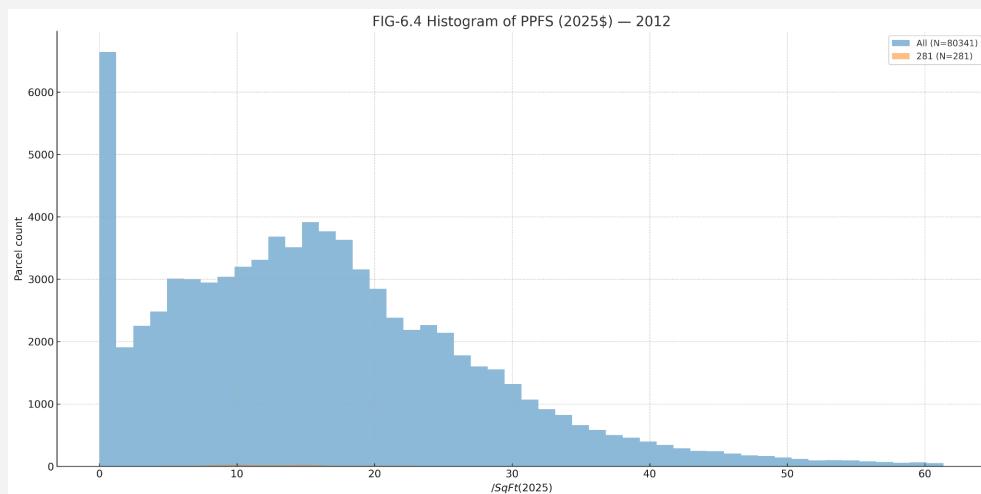


Figure 6.4 — Histograms (2025\$; All vs 281) — 2012

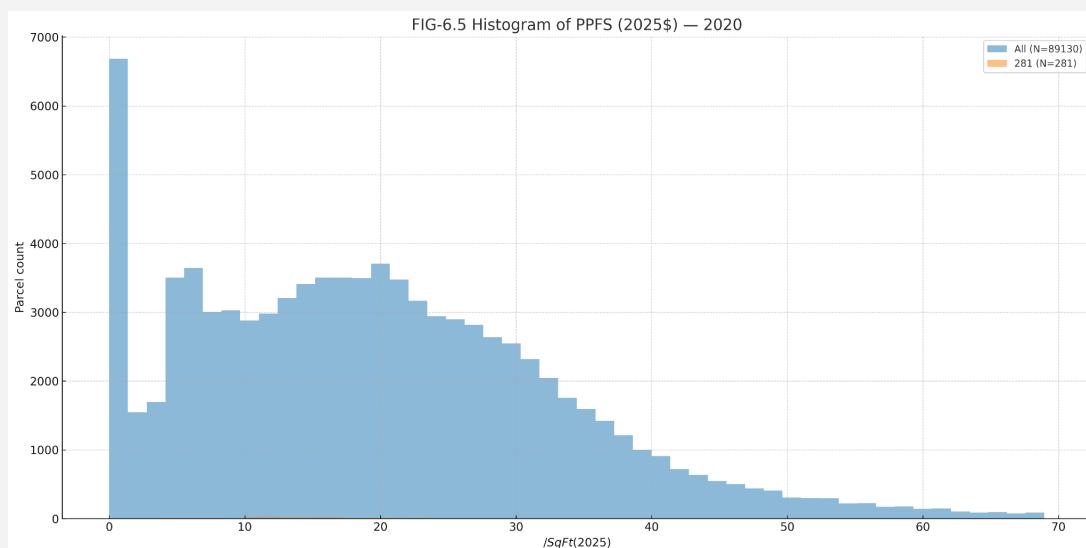


Figure 6.5 — Histograms (2025\$; All vs 281) — 2020

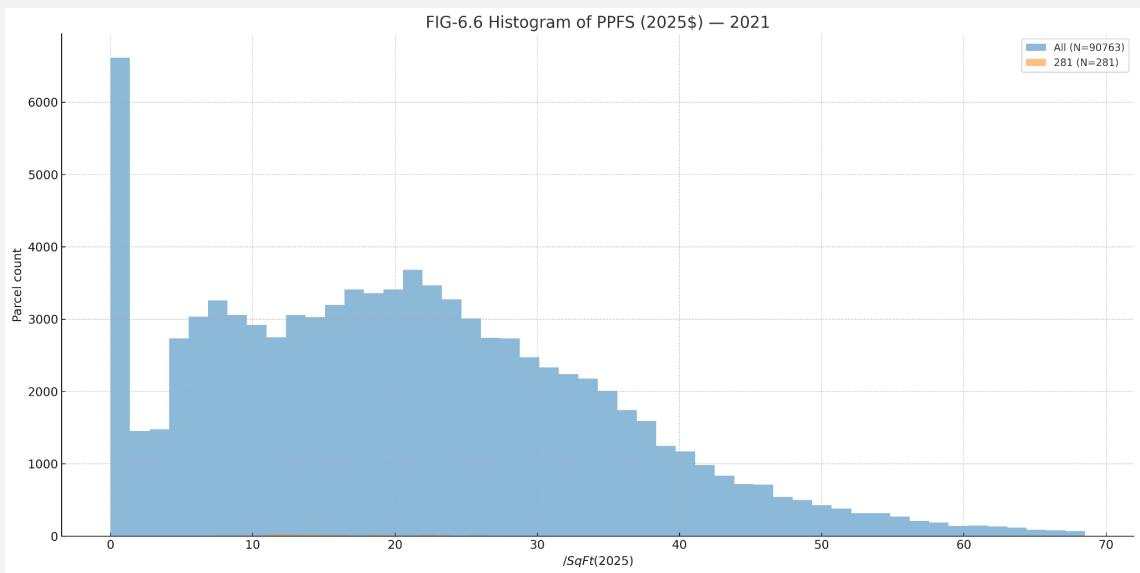


Figure 6.6 — Histograms (2025\$; All vs 281) — 2021

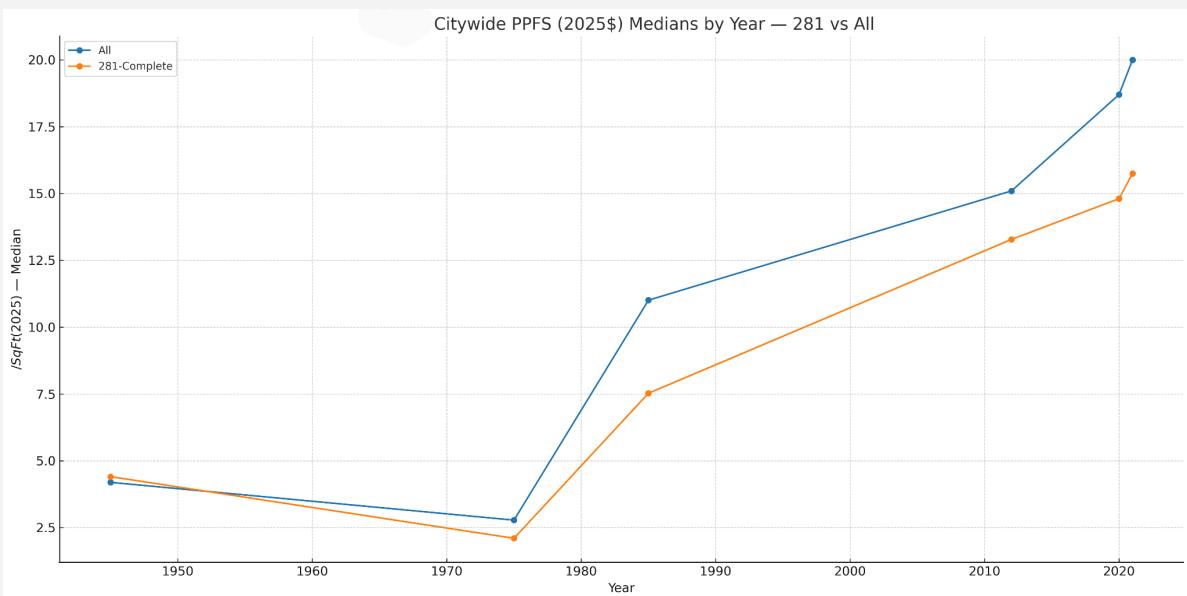


Figure 6.7 — Citywide PPFS (2025\$) Median by Year — 281 vs All

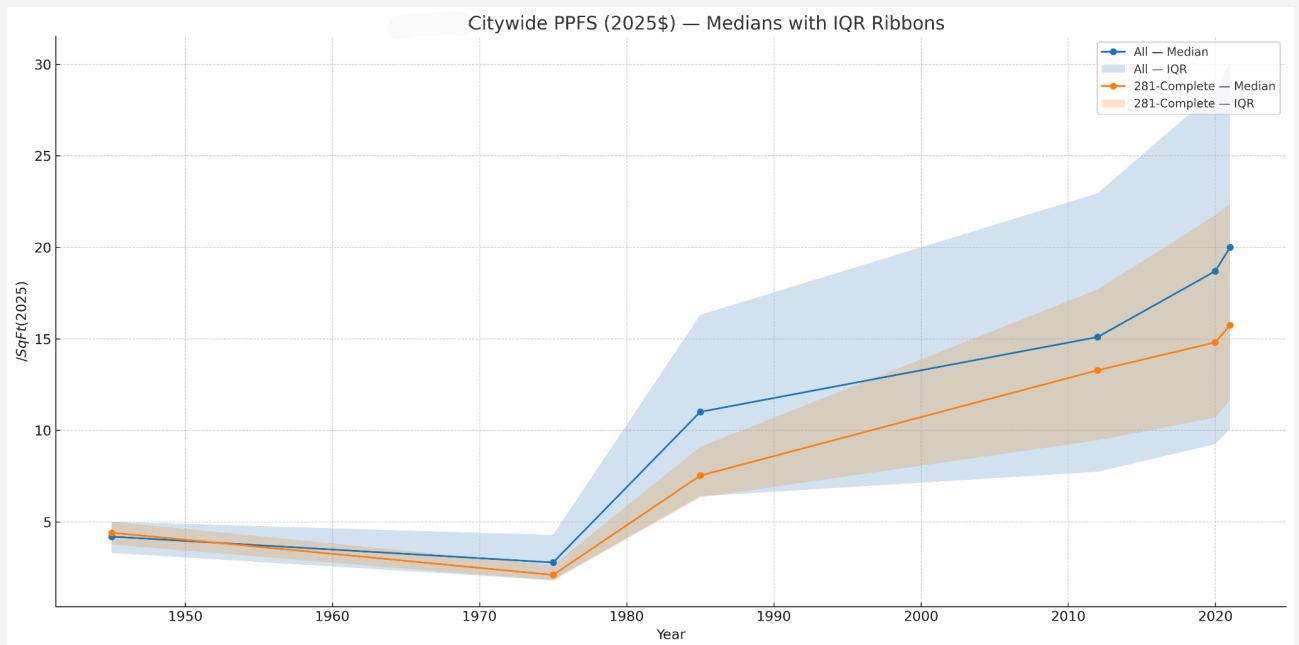


Figure 6.8 — Citywide PPFS (2025\$) Median with IQR Ribbons

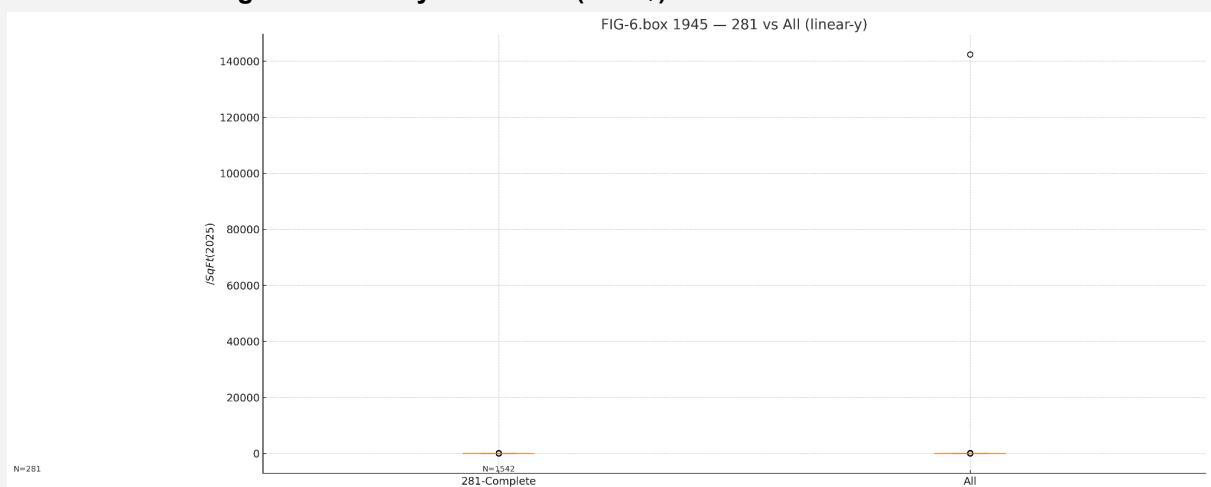


Figure 6.9 — Boxplots (linear & log): 1945 L

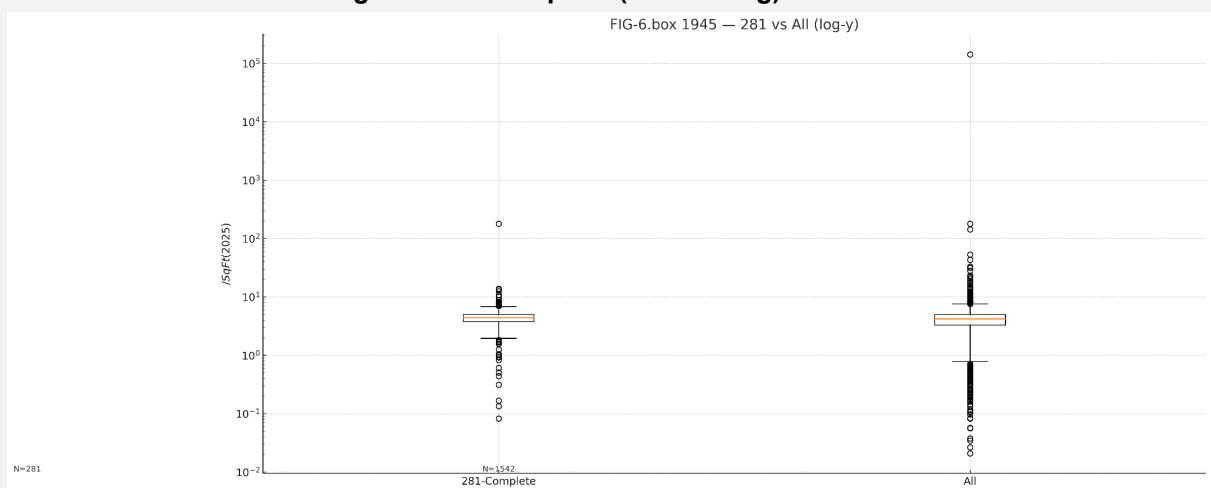


Figure 6.10 — Boxplots (linear & log): 1945 Log

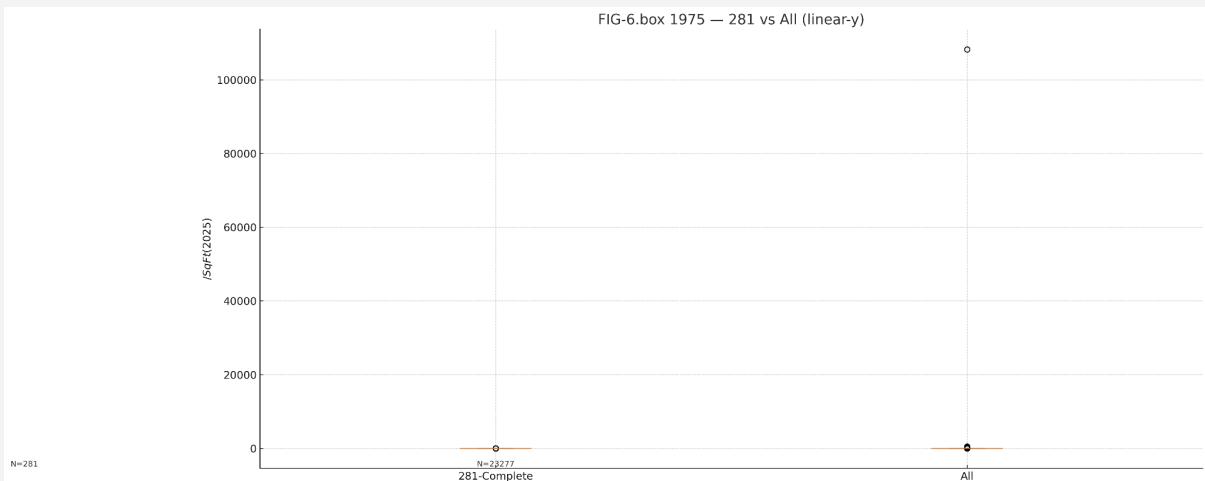


Figure 6.11 — Boxplots (linear & log): 1975 L

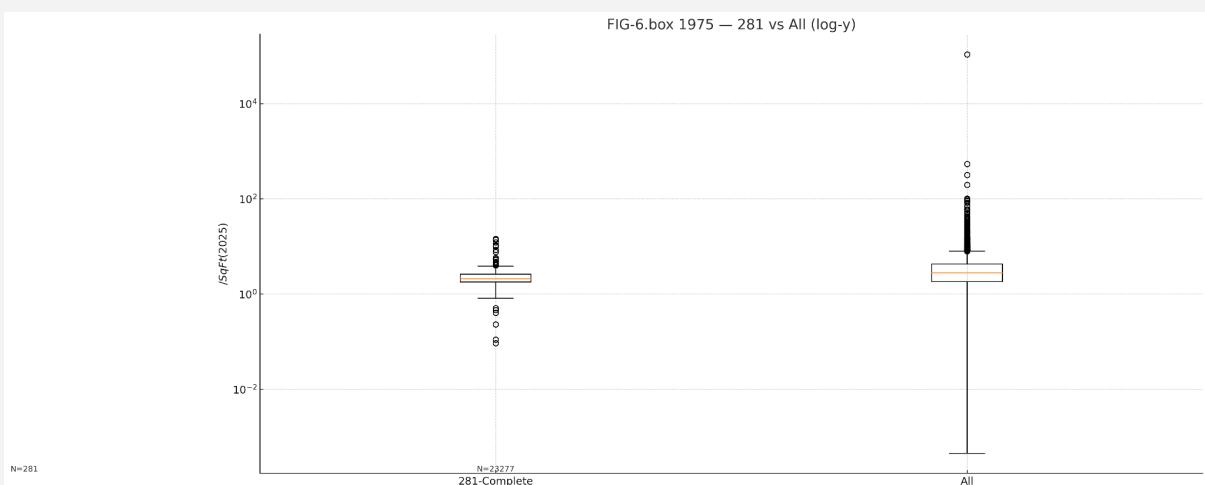


Figure 6.12 — Boxplots (linear & log): 1975 Log

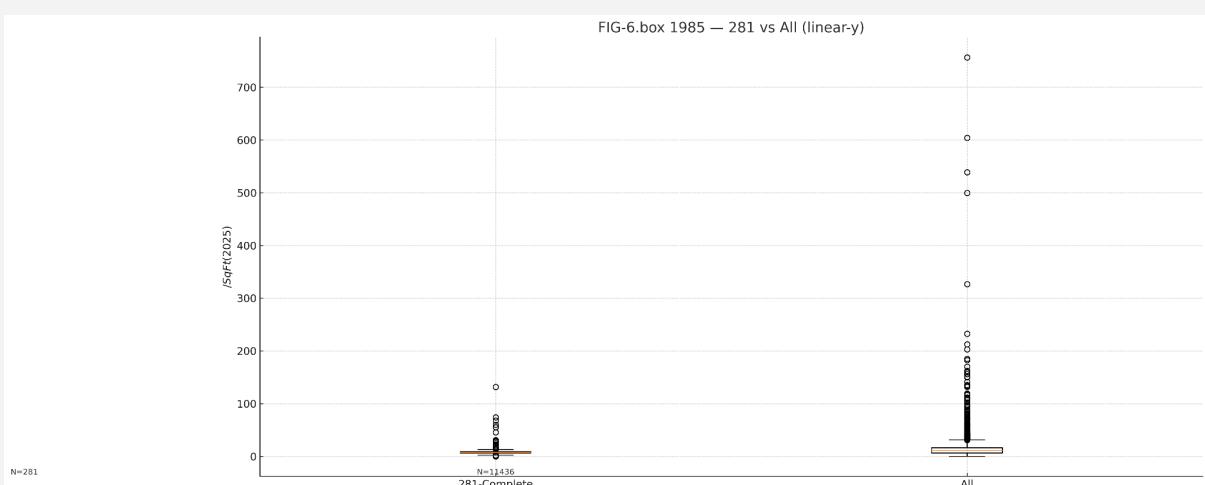


Figure 6.13 — Boxplots (linear & log): 1985 L

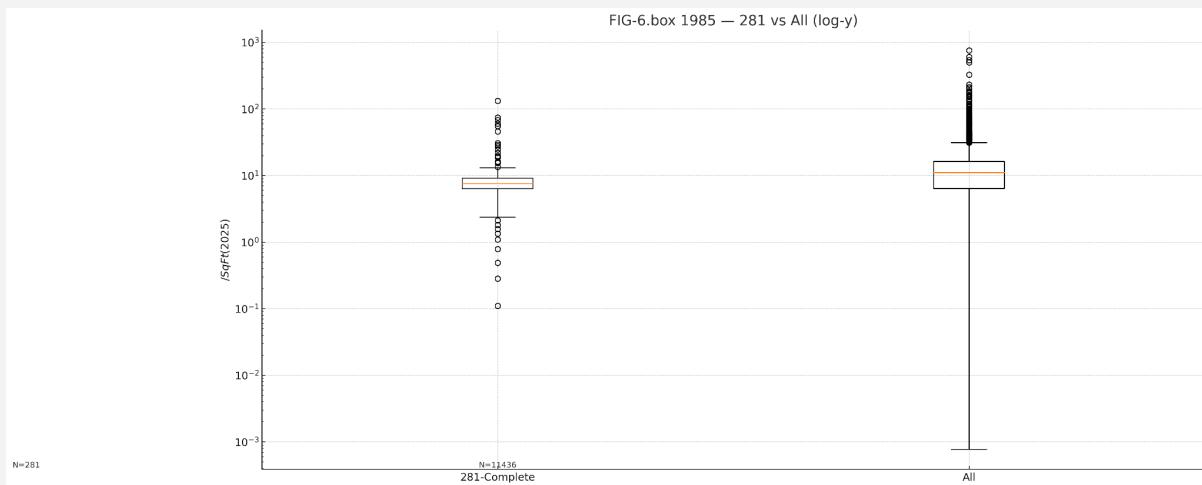


Figure 6.14 — Boxplots (linear & log): 1985 Log

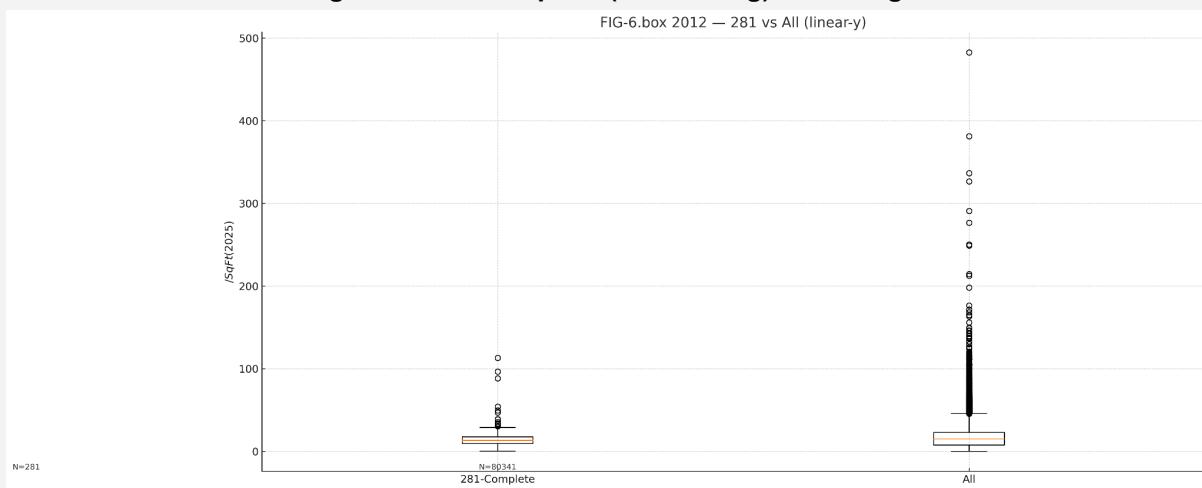


Figure 6.15 — Boxplots (linear & log): 2012 L

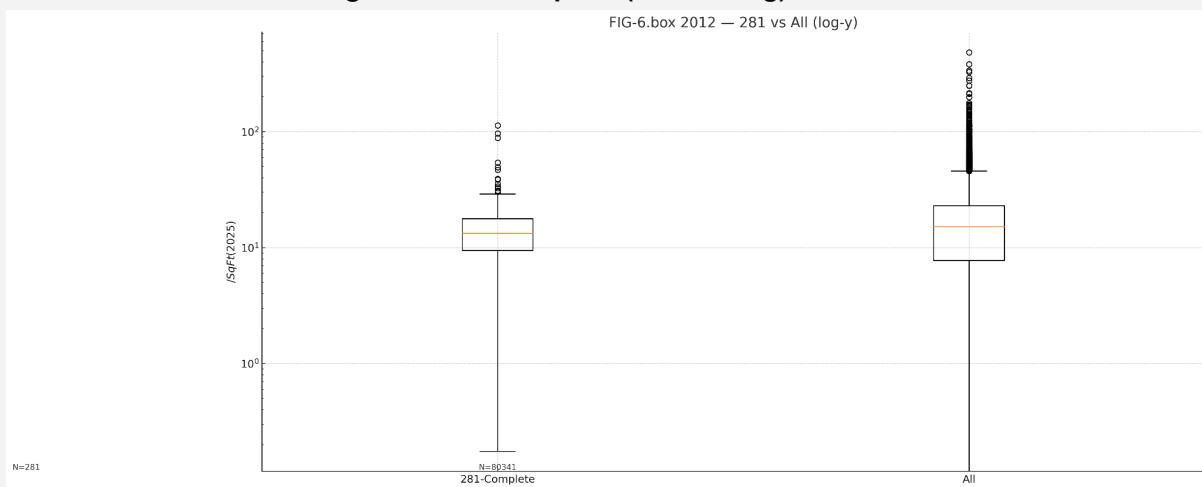


Figure 6.16 — Boxplots (linear & log): 2012 Log

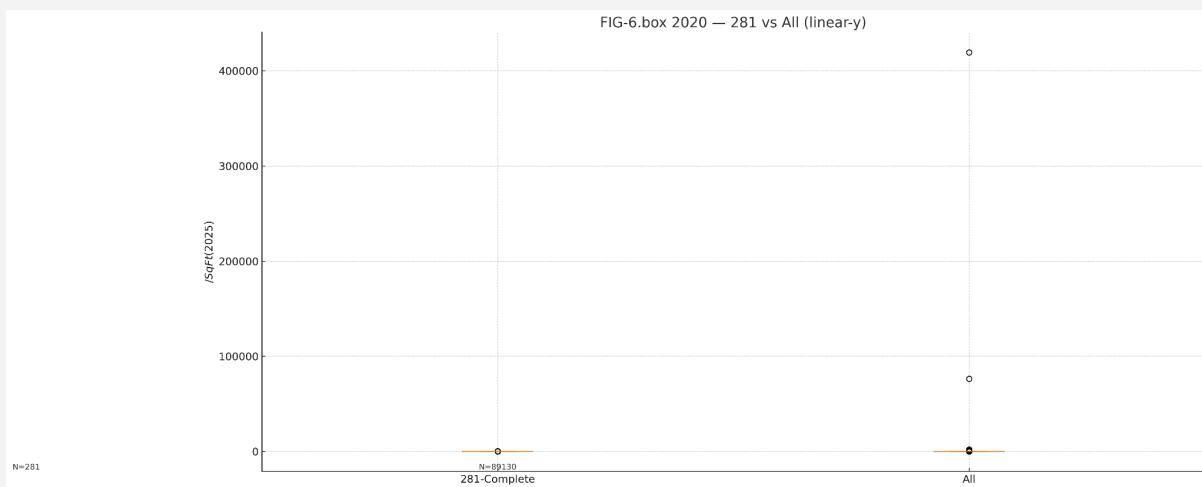


Figure 6.17 — Boxplots (linear & log): 2020 L

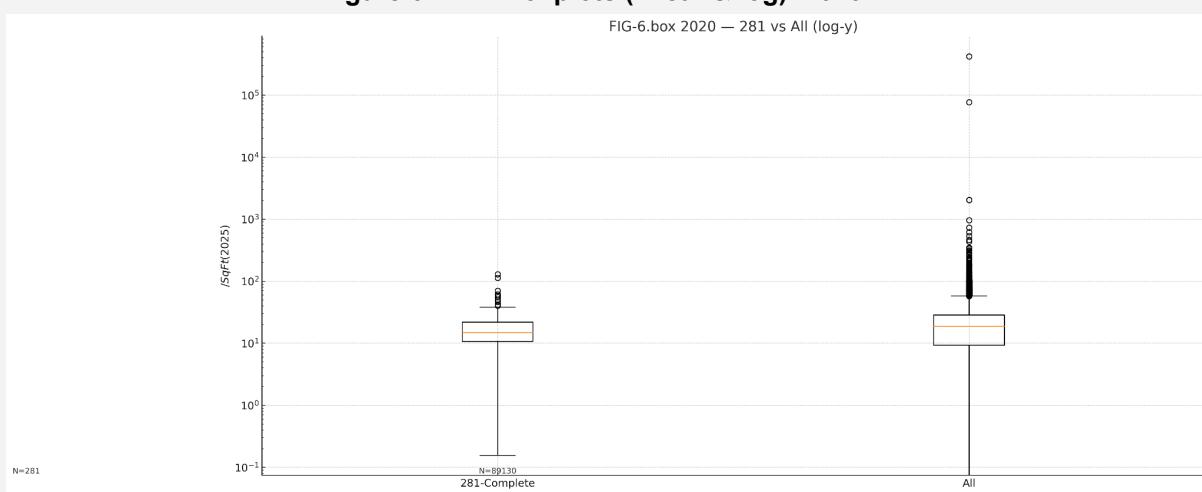
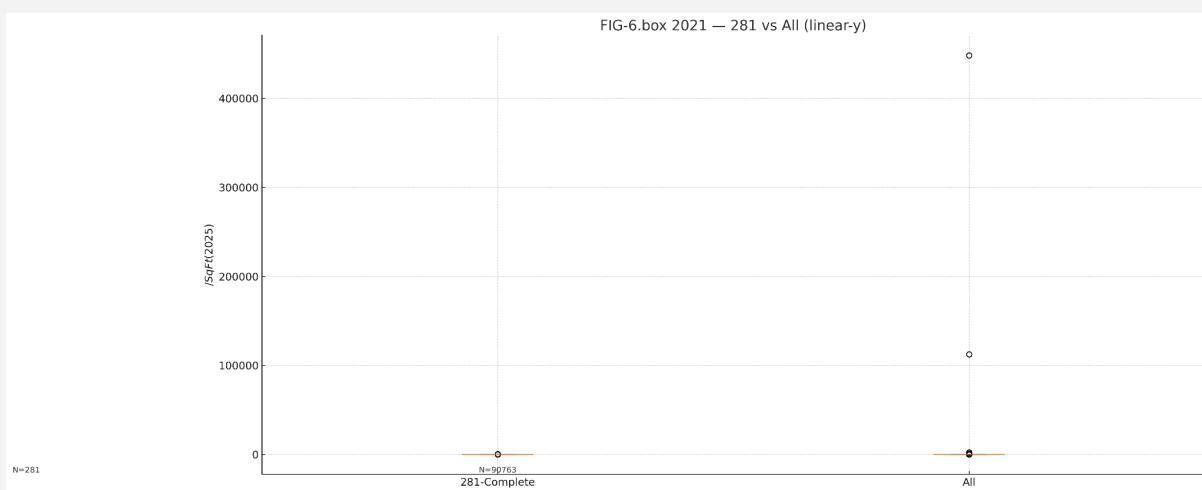
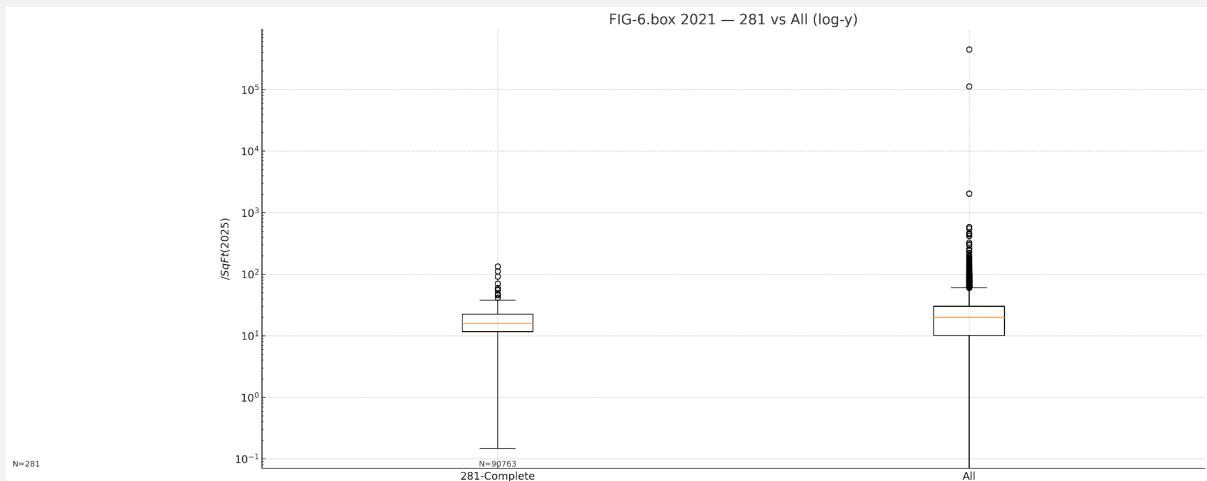


Figure 6.18 — Boxplots (linear & log): 2020 Log



Figures 6.19 — Boxplots (linear & log): 2021 L



Figures 6.20 — Boxplots (linear & log): 2021 Log

6.4 Nominal \$ Benchmarks (single-year interpretation)

Nominal medians provide within-year context (inflation not applied):

- All parcels: 1945 = \$0.24, 1975 = \$0.47, 1985 = \$3.76, 2012 = \$11.02, 2020 = \$15.33, 2021 = \$17.24 /sqft.
 - Complete (281): 1945 = \$0.25, 1975 = \$0.36, 1985 = \$2.57, 2012 = \$9.69, 2020 = \$12.13, 2021 = \$13.57 /sqft.
- Use nominal for cross-sectional reading; use 2025\$ when comparing across years.

Year	Cohort	N	Median
1945	281-Complete	281	0.251177394
1945	All	1542	0.239032621
1975	281-Complete	281	0.357142857
1975	All	23277	0.47293218
1985	281-Complete	281	2.567714286
1985	All	11436	3.756207774
2012	281-Complete	281	9.692333333
2012	All	80341	11.01723383
2020	281-Complete	281	12.13371429
2020	All	89130	15.32652237
2021	281-Complete	281	13.57085714
2021	All	90763	17.23624511

Table 6.3 — Nominal PPFS Medians by Year and Cohort

6.5 Distribution Shape and Tail Diagnostics

Histograms (FIG-6.1..6.6) and log-scaled boxplots (FIG-6.9..6.20) show consistent right-skew with a thin but influential upper tail, particularly in All. A key mechanical driver is **very small denominators** (micro-parcels) producing high \$/sqft; hence our reliance on medians/IQR and the inclusion of log views. CCDFs in §6.9 further clarify top-end behavior.

6.6 Zone Gradient — 2021 (2025\$)

We compare typical PPFS-2025\$ across **Redlining core, Half-mile, 0.5–1 mile, Outside-Both** for All and **281**. Bars are double-coded with N above each bar.

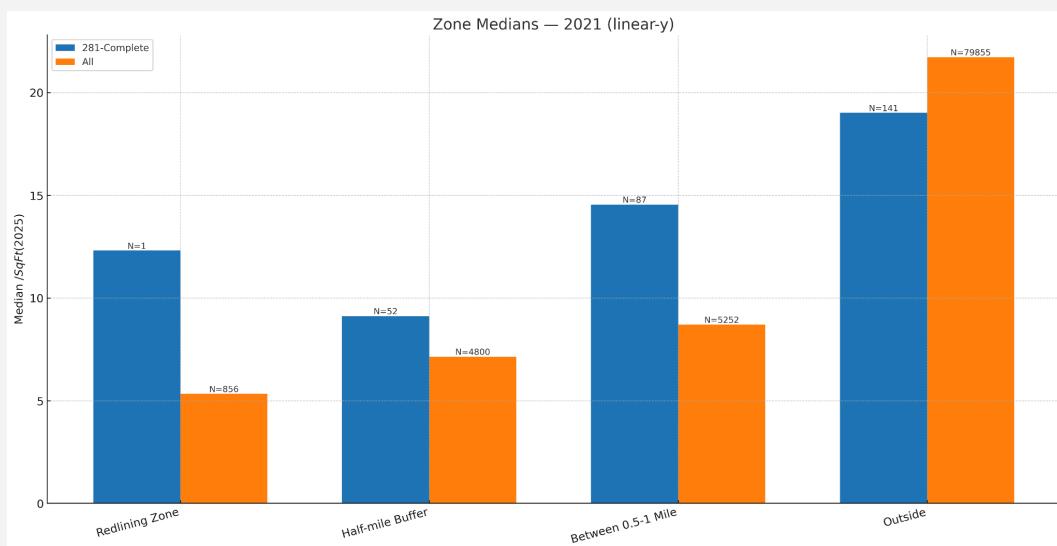


Figure 6.21 — Zone medians 2021 (linear)

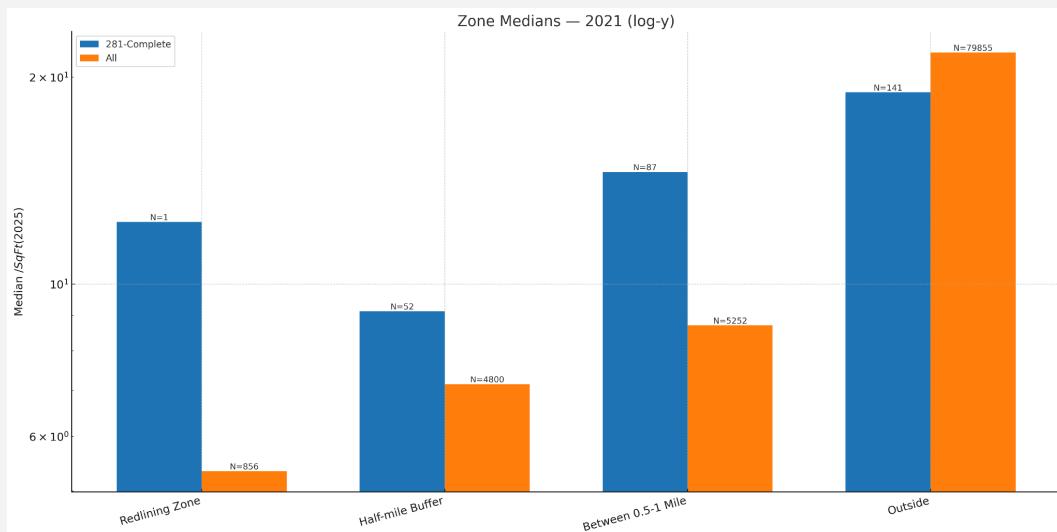


Figure 6.22 — Zone medians 2021 (log)

Trend view by zone. To test persistence, we track medians across all six years for each zone (All vs 281 in parallel):

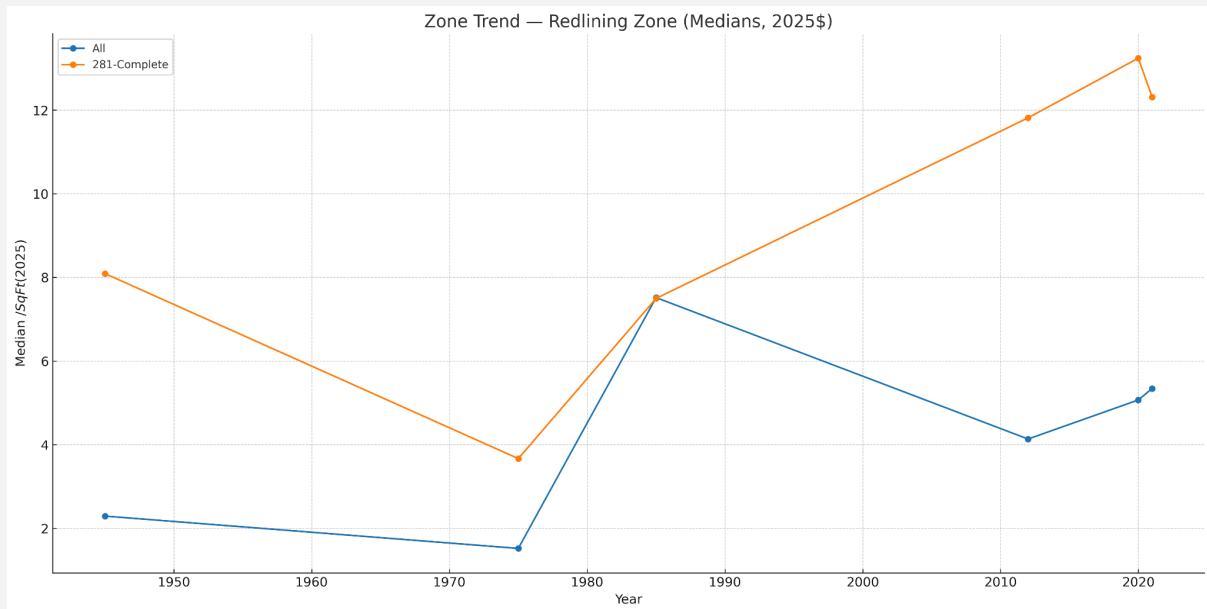


Figure 6.23 — Redlining Zone — trend of medians across years

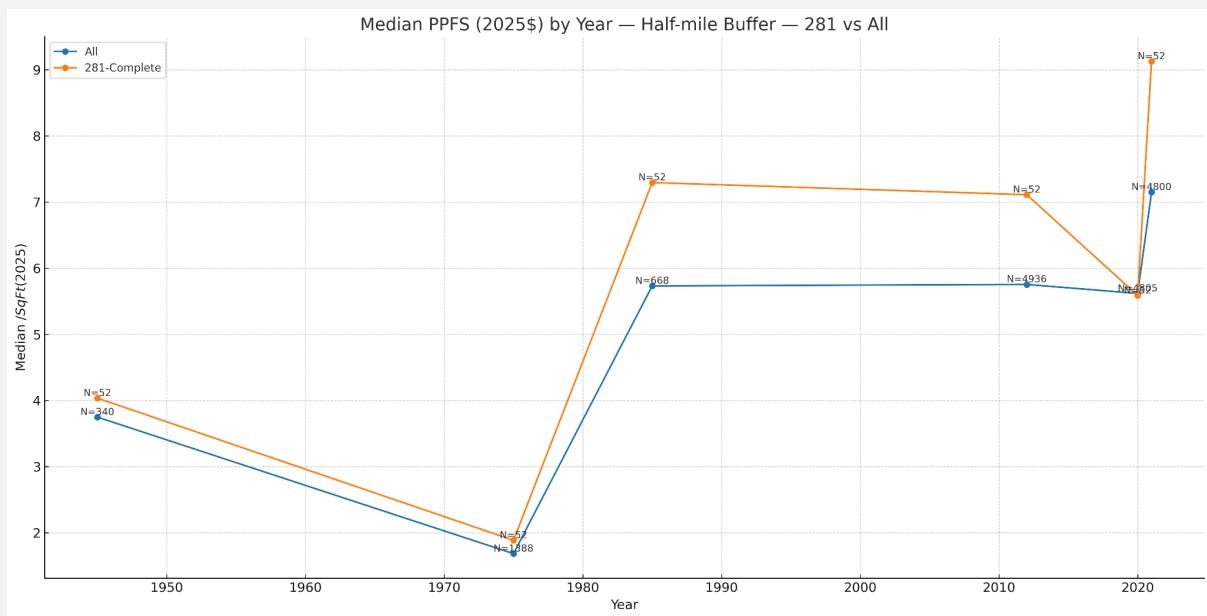


Figure 6.24 — Half-mile Buffer — trend of medians across years

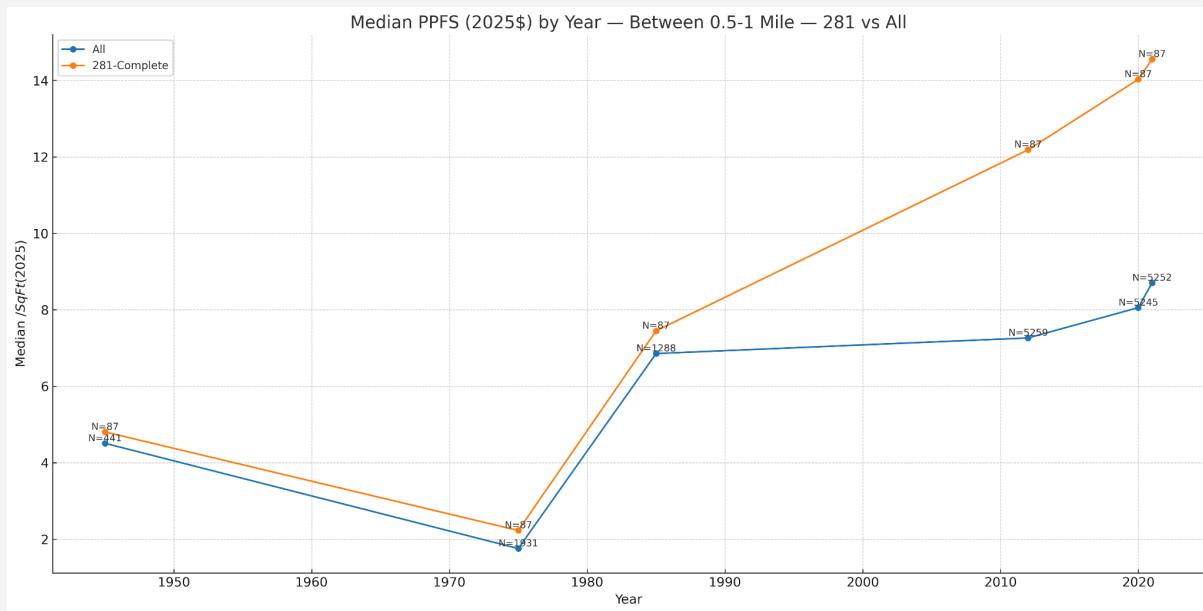


Figure 6.25 — Between 0.5–1 mile — trend of medians across years

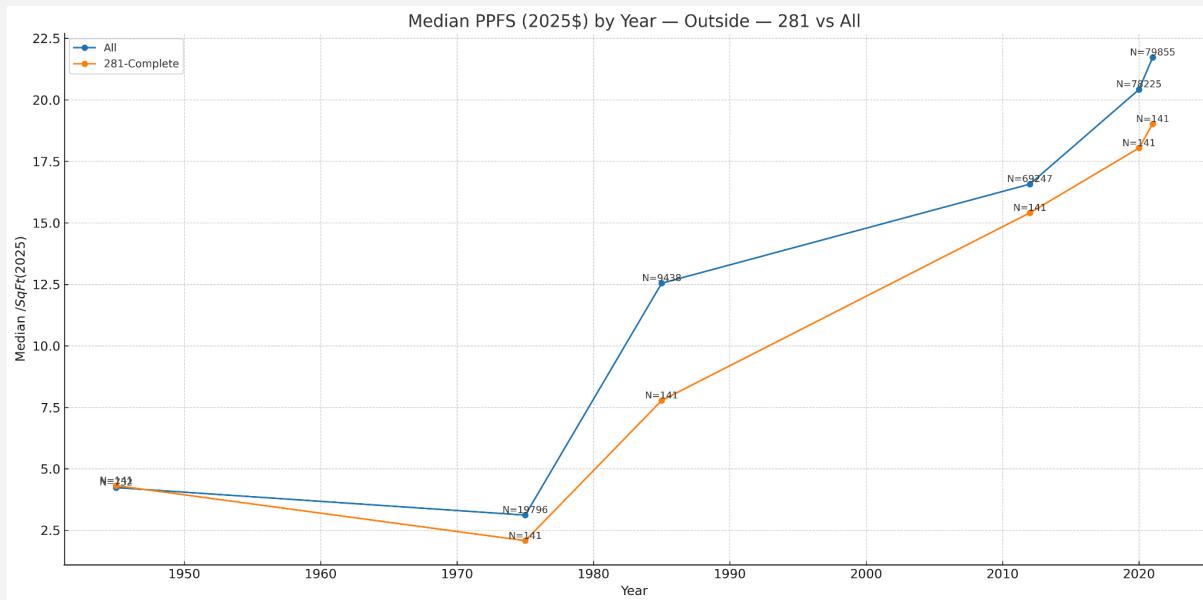


Figure 6.26 — Outside Both — trend of medians across years

6.7 PPFS Growth

We evaluate growth as **log differences** of PPFS-2025\$ between adjacent pairs (1945→1975, 1975→1985, 2012→2020, 2020→2021) and a longer arc (1985→2021). Within-address results use only the 281 cohort; citywide results use a **pseudo-panel** (log change of citywide medians). Both are annotated with Ns.

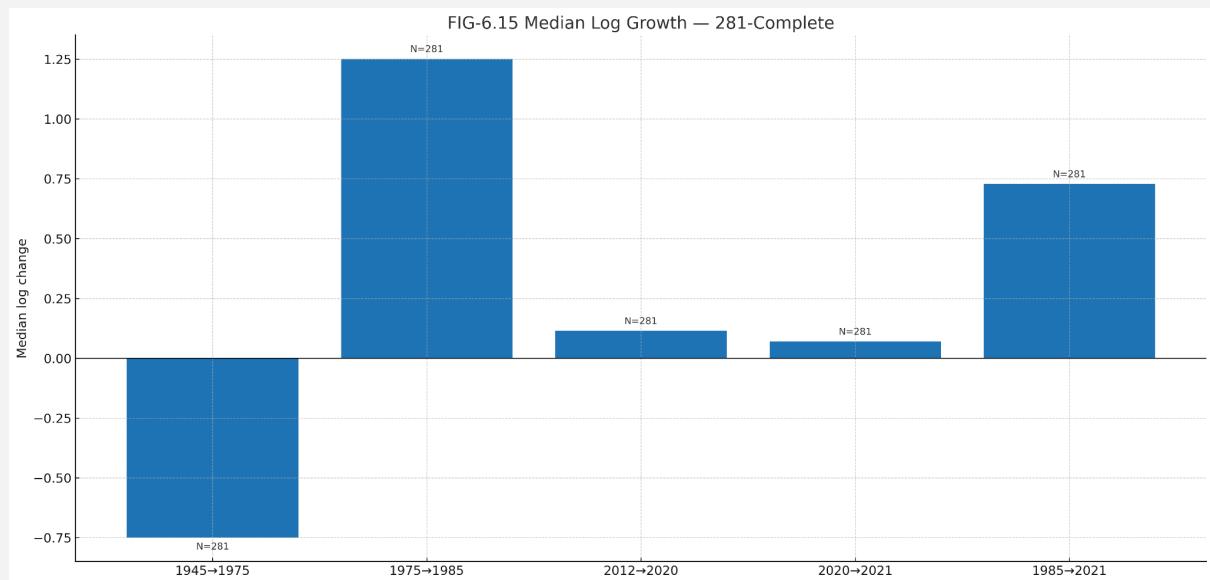


Figure 6.27 — Median log growth — 281 (within-address)

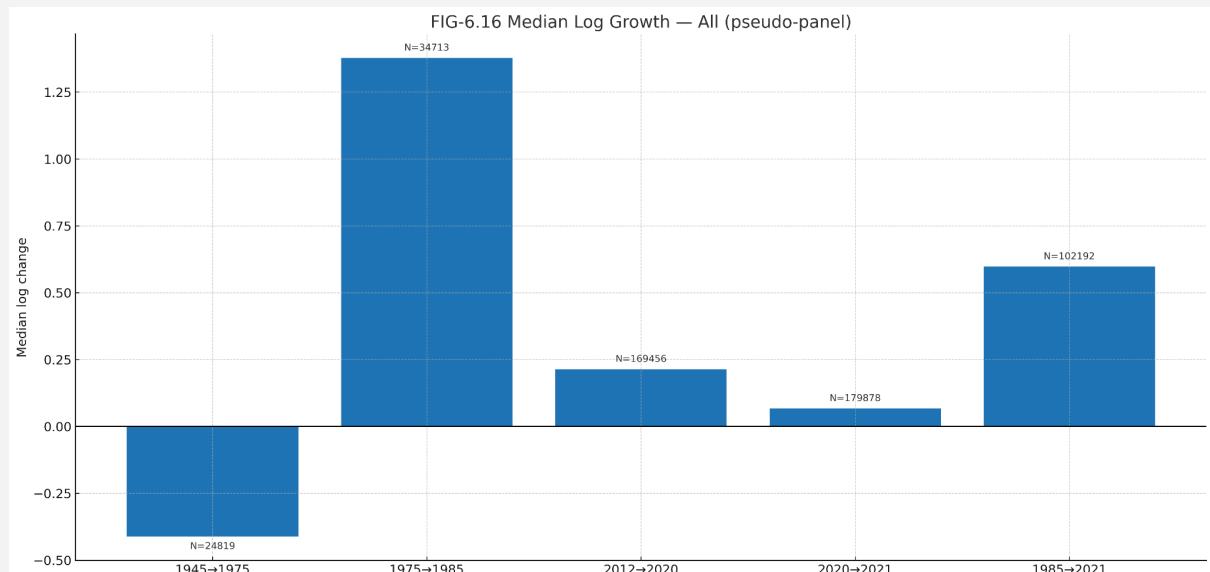


Figure 6.28 — Median log growth — All (pseudo-panel)

6.8 Additional Visual Diagnostics

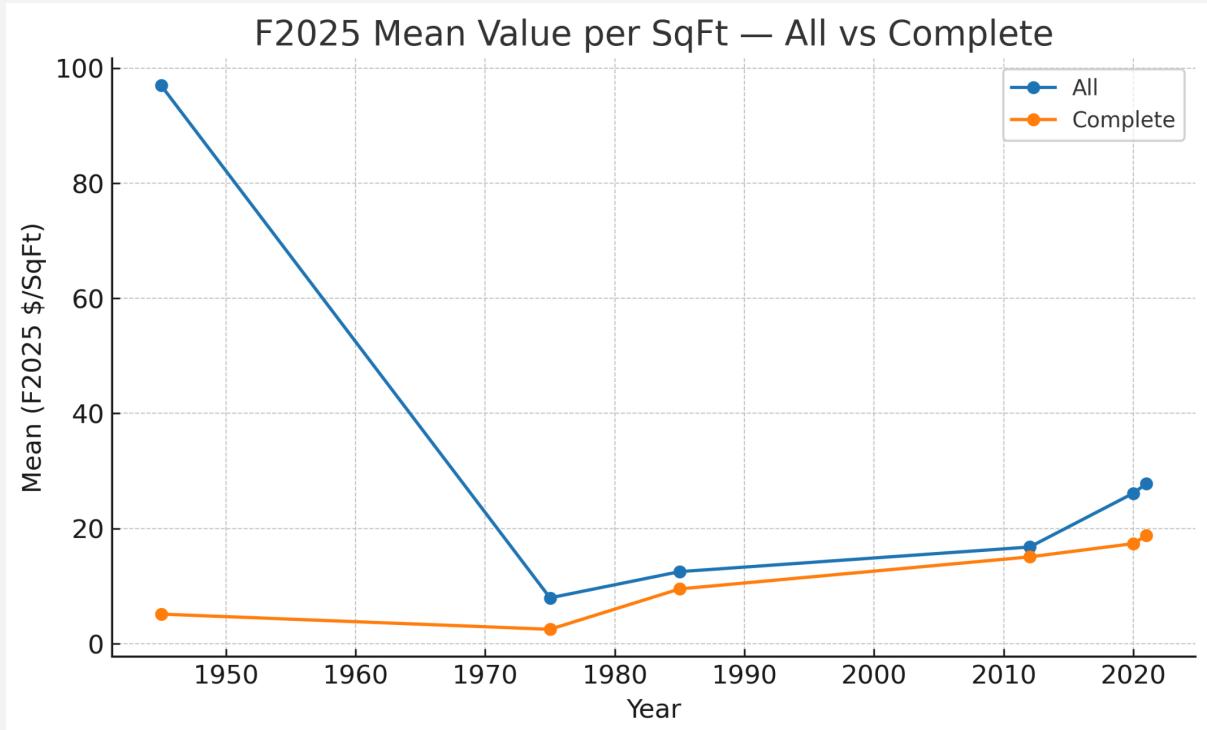


Figure 6.29 — F2025 Mean Value per SqFt - All vs Complete

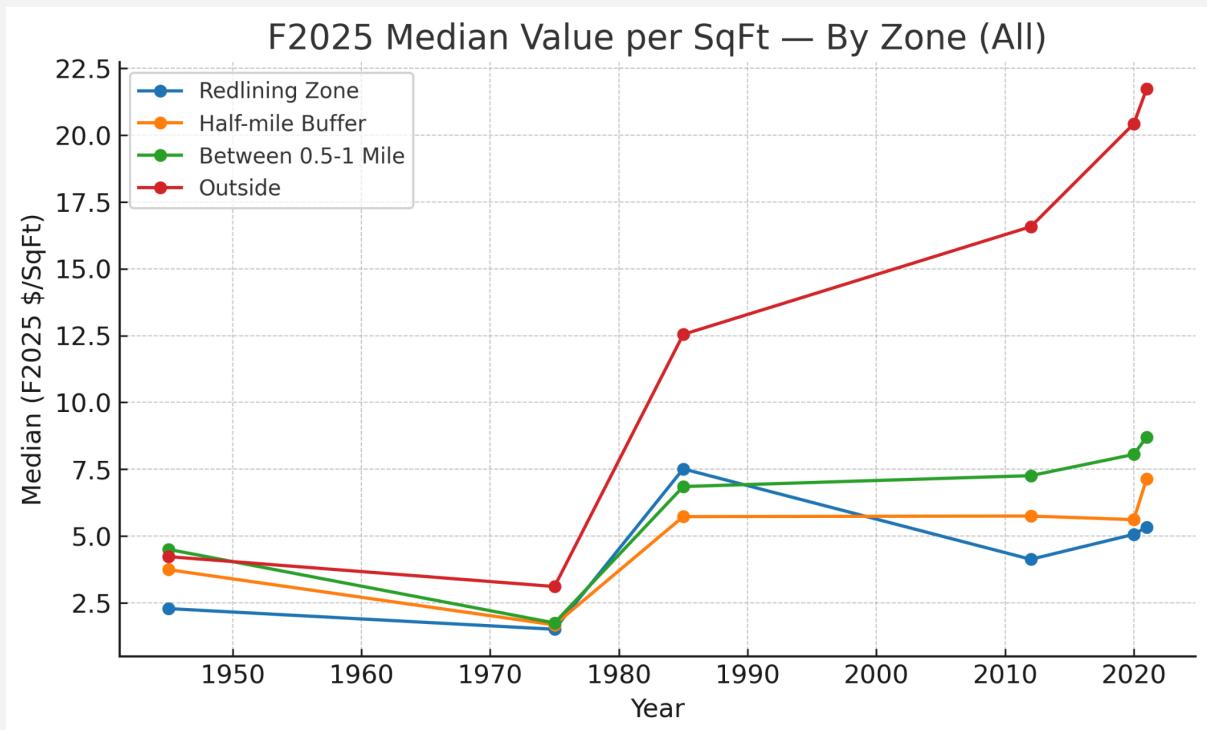


Figure 6.30 — F2025 Mean Value per SqFt - By Zone (All)

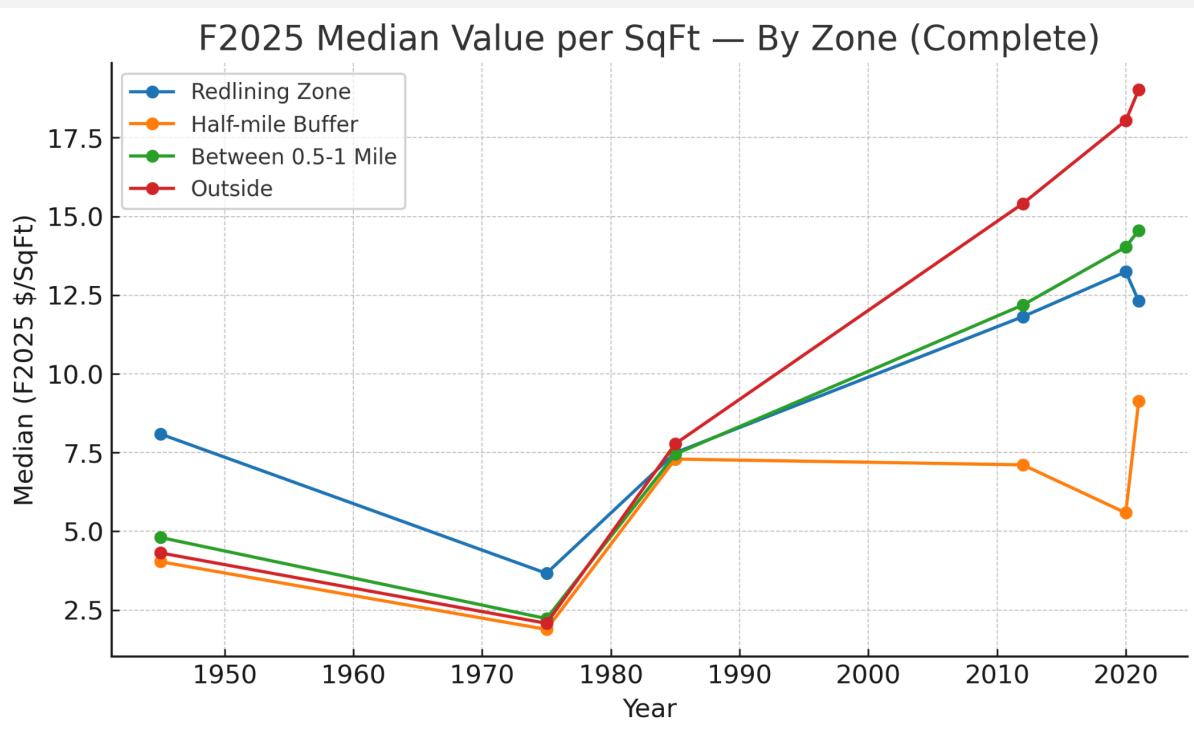


Figure 6.31 — F2025 Mean Value per SqFt - By Zone (Complete)

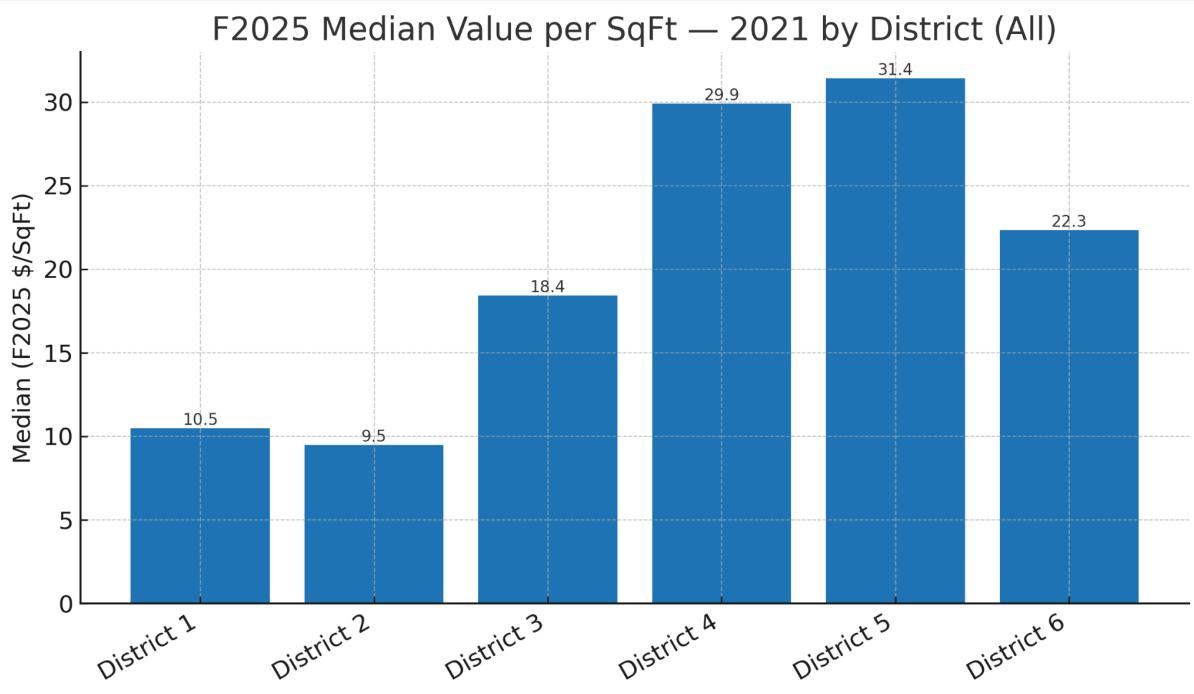


Figure 6.32 — F2025 Mean Value per SqFt - 2021 District (All)

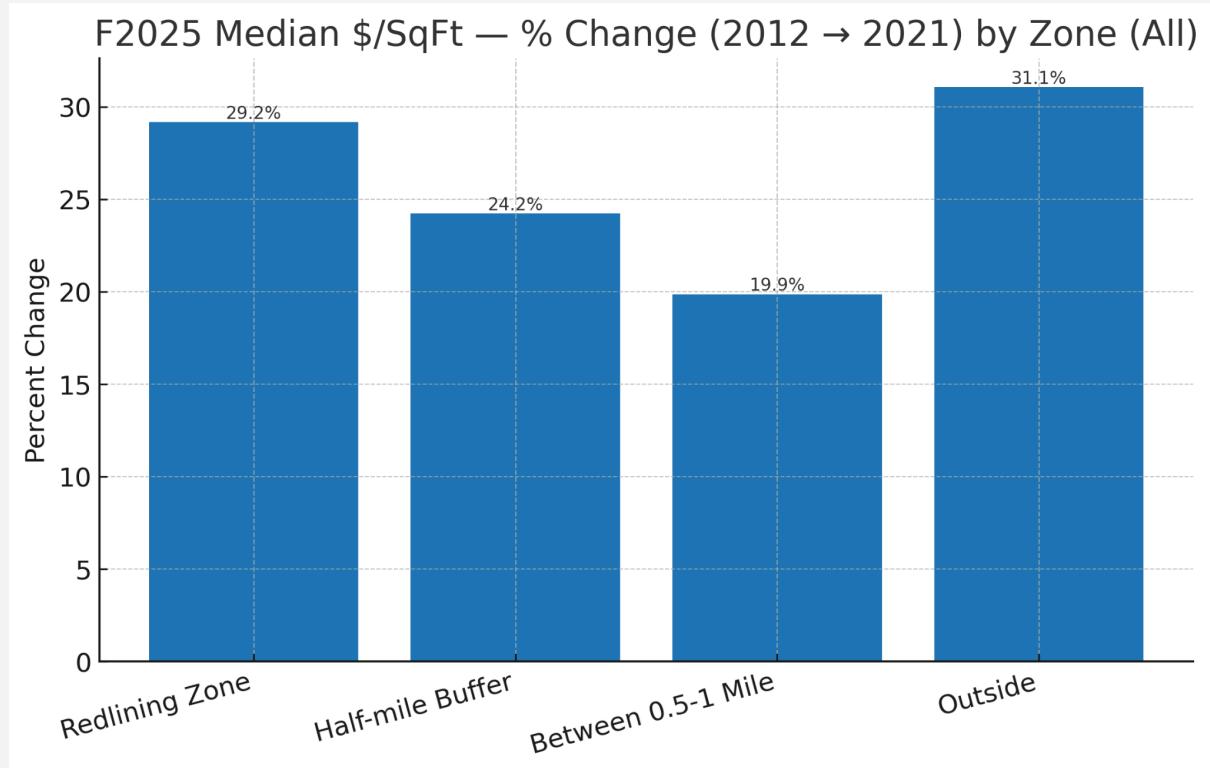


Figure 6.33 — F2025 Median \$/Sqft - %Change (2012→2021) by Zone (All)

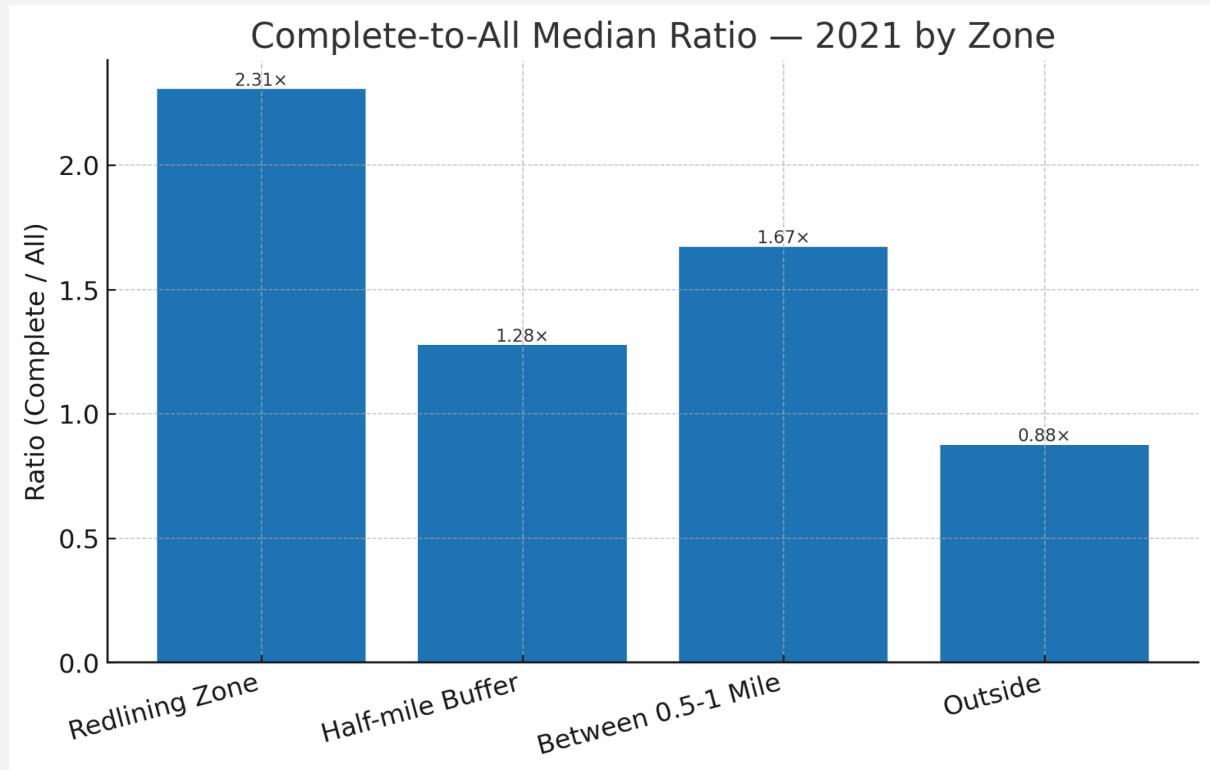


Figure 6.34 — Complete-to-All Median Ratio - 2021 by Zone

6.8.1 What these figures show

- **All vs Complete (Means).** Citywide F2025 means rise across the six benchmarks for both groups; however, the “All” means are inflated by heavy right tails (e.g., **1945 mean 96.98** vs **median 4.19** \$2025/sqft in your table). This gap is persistent, so **medians** are the safer summary for policy interpretation.
- **Zone time trends (All).** The **Outside** ring leads at every date; core-adjacent areas trail:
 - 2025\$ zone medians (All) move from **2012→2021** by:
 - **Redlining Zone:** 4.13 → **5.34 (+29%)**
 - **Half-mile:** 5.75 → **7.15 (+24%)**
 - **0.5–1 mile:** 7.26 → **8.71 (+20%)**
 - **Outside:** 16.57 → **21.72 (+31%)**
 - Gains are broad-based, with the largest percentage rise **outside** the buffers.
- **Zone time trends (Complete).** The like-for-like cohort tracks similar ordering but sits **higher** than “All” in the inner zones and **lower** outside:
 - **Complete/All median ratio in 2021:**
 - Redlining **2.31×** (12.31 vs 5.34)
 - Half-mile **1.28×** (9.13 vs 7.15)
 - 0.5–1 mile **1.67×** (14.56 vs 8.71)
 - Outside **0.88×** (19.02 vs 21.72)
 - This pattern is consistent with selection into the 281-parcel panel (more stable core parcels retained through time; outer-ring coverage is broader in “All”).
- **District cross-section (2021, All).** Median PPFS (2025\$) is highest in **District 5 (31.42)** and **District 4 (29.89)**, mid in **District 6 (22.33)** and **District 3 (18.43)**, and lowest in **District 1 (10.49)** and **District 2 (9.47)** — mirroring earlier sections’ rankings and the strong outer-ring gradient.

6.8.2 Conclusions

- **Citywide appreciation** (2012→2021) is broad but **concentrated at the edge**: Outside parcels posted the largest median gains (~+31%), with core and buffer zones rising +20–29%.
- **Persistent spatial gradient:** In every modern benchmark, **Outside** medians exceed buffer and core medians; 2021 medians continue to rank **Redlining < Half-mile < 0.5–1 mile < Outside**.
- **Panel selection matters:** The **Complete (281)** cohort shows **higher medians** in the inner zones than the “All” sample (e.g., 2021 Redlining **12.31** vs **5.34**), but a **lower** outside-ring median (**19.02** vs **21.72**). Balanced-panel findings should be read as **structural** (less influenced by churn/micro-lots), while “All” captures **full cross-section** heterogeneity including outliers.
- **District inequality** in 2021 is large: **D5/D4** lead with ~30–31 \$/sqft medians; **D1/D2** trail near 9–10.5 \$/sqft. Within-district spreads are modest relative to **between-district** differences.
- **Means vs medians:** Extremely high maxima in the “All” dataset (esp. early years) keep means well above medians; **medians/IQRs** remain the recommended standard for comparisons.

7. Persistent Effects of Jim Crow-Era Redlining on Lubbock Property Values

7.1 Era-Appropriate City Limits: Source, Rule, and Integration

City incorporation status conditions access to public goods and administrative regimes that shape assessed values. To separate redlining effects from cityhood effects, each parcel receives a **study-year-specific city-limits status** based on era-appropriate municipal boundaries.

Source: City of Lubbock *Annexation History by Decade* map, providing decade-aggregated polygons for 1900s, 1920s, 1930s, 1940s, 1950s, 1970s, 1980s, 1990s, 2000s, 2010s, and 2020s (no standalone 1960s layer).

Classification rule (before, not after). For study year t , the boundary used is the **latest decade that fully predates t** . This avoids forward-looking contamination from annexations occurring later in the same decade.

Study year	Boundary decade used	Layer name
1945	1940s (pre-1945)	city_limit_1945
1975	1950s (pre-1975; 1960s not available)	city_limit_1975
1985	1970s (pre-1985)	city_limit_1985
2012	2010s (pre-2012)	city_limit_2012
2020	2010s (pre-2020)	city_limit_2020
2021	2020s (pre-2021)	city_limit_2021

Table 7.1 — Mapping of Study Year to City Limit Boundary Decade

Why not 1970s for 1975 or 1980s for 1985? Those decade polygons include future-year annexations (1976–1979; 1986–1989). Using them would leak information from the future. The conservative choice is to use the last **fully prior** decade (1950s for 1975; 1970s for 1985).

Georeferencing and preparation (ArcGIS Pro). The PDF was georeferenced to the project CRS; decade polygons were traced/dissolved into single features, verified for nesting across decades, and exported as the six study-year layers above.

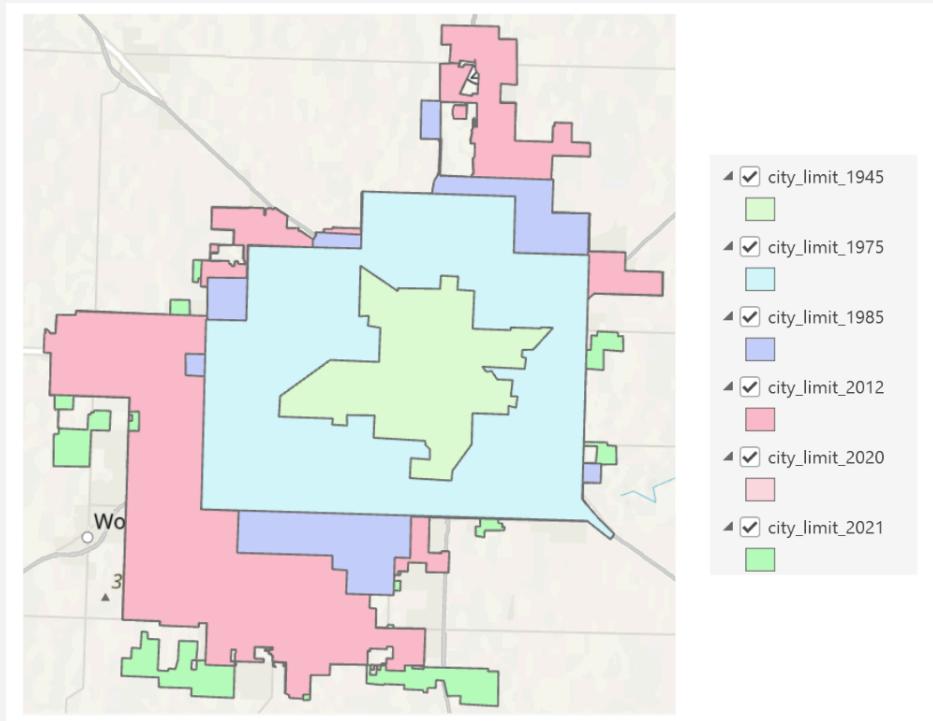


Figure 7.1 — city_limit_1945, city_limit_1975, city_limit_1985, city_limit_2012, city_limit_2020, city_limit_2021

7.2 Possible Redlining Zones (RL1, RL2, RL3)

Define alternative, evidence-based operationalizations of a redlining footprint to test robustness of results to boundary choice. Three candidate polygons are used: RL1 (1923 ordinance-based construction), RL2 (highway-era construction), and RL3 (expert-provided polygon). Each has three non-overlapping ring buffers at 0–0.5 mi (`_buf1`), 0.5–1.0 mi (`_buf2`), and 1.0–1.5 mi (`_buf3`).

7.2.1 RL1 (1923 ordinance geometry)

A rectangle formed by extending Avenue C (north–south) and 16th Street (east–west) to an intersection, then adding one parallel line east of Avenue C and one parallel line south of 16th Street so that the southeast portion of the 2021 city limit fits entirely within the polygon. Parallel lines were drawn to intersect and close the polygon; specific offsets were not measured in physical units. Layer names: `RL1`, `RL1_buf1`, `RL1_buf2`, `RL1_buf3`.

7.2.2 RL2 (highway-era geometry)

Constructed analogously to RL1, substituting Interstate 27 for Avenue C while retaining 16th Street as the east–west axis. Parallel lines were drawn to form a closed rectangle under the same design principle as RL1. This construction reflects reporting that Interstate 27 now functions as the barrier historically associated with Avenue C (Texas Tribune, 2023; 2024). Layer names: `RL2`, `RL2_buf1`, `RL2_buf2`, `RL2_buf3`.

7.2.3 RL3 (expert consultation)

A polygon supplied via expert consultation and used as provided. Layer names: `RL3`, `RL3_buf1`, `RL3_buf2`, `RL3_buf3`.

7.2.4 Buffers for RL1–RL3

For each RL core polygon, three bands are created as donut rings at 0–0.5 mi, 0.5–1.0 mi, and 1.0–1.5 mi measured from the polygon boundary; band layers follow the `_buf1`, `_buf2`, `_buf3` naming pattern shown above.

Inferential analyses are executed separately for each RL variant (RL1-only, RL2-only, RL3-only). When multiple RL variants are displayed simultaneously for exploratory mapping, parcel overlaps can occur but are not used for inferential summaries.

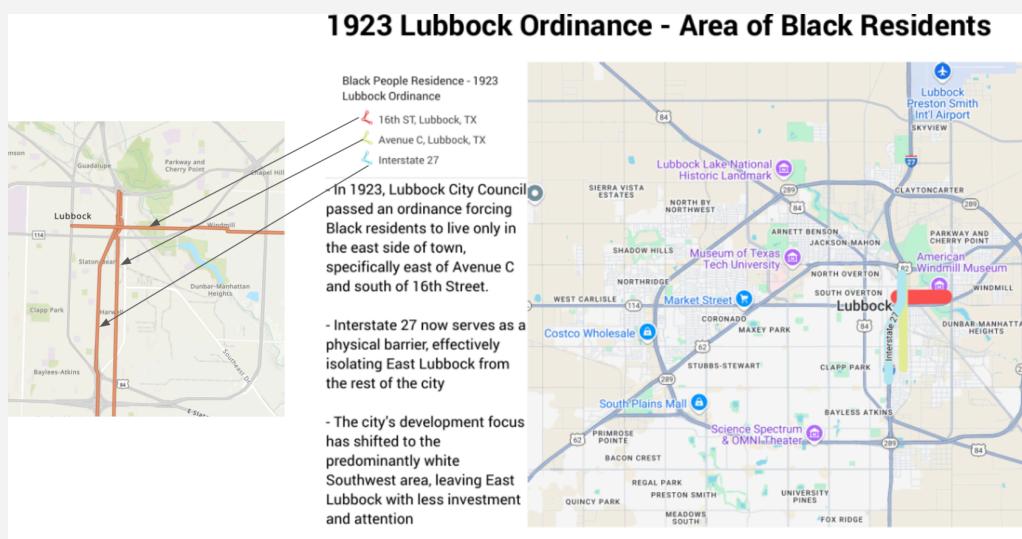


Figure 7.2 — Lubbock ordinance context and road references



Figure 7.3 — Possible redlining zones considered (RL1, RL2, RL3 and buffers)

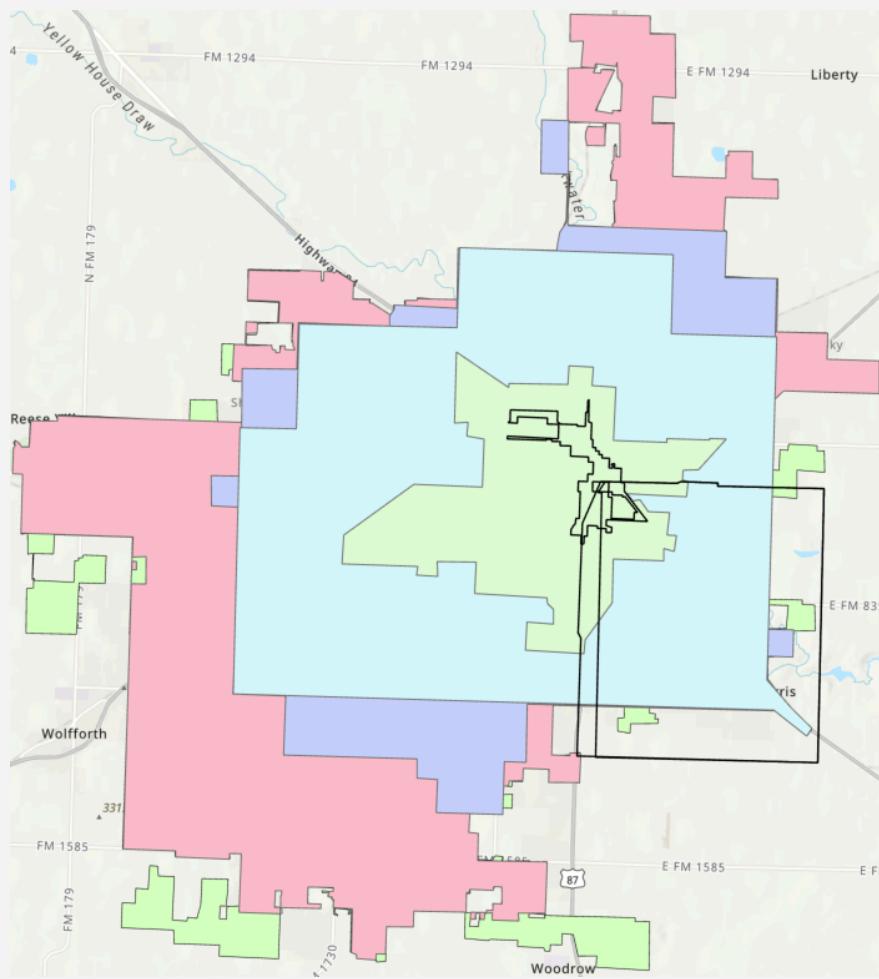


Figure 7.4 — All City Limits and Redlining Zones together

7.3 Analytical Framework and Cohorts

7.3.1 Inputs and provenance

The analysis relies on a single, consolidated parcel-level dataset, a companion Python script that operationalizes the transformations described earlier in the report, and the era-specific municipal boundary and redlining-zone layers prepared in §§7.1–7.2.

The **primary dataset** contains the parcel-level attributes required for constructing the outcomes used throughout Section 7 (assessed value totals and per-square-foot variants in both nominal and CPI-adjusted dollars), for linking parcels to study years (1945, 1975, 1985, 2012, 2020, 2021), and for joining to spatial classifications. Field names follow the conventions laid out in §7.3.2 below and reflect the naming already used in the analysis code.

The **analysis script** reads the CSV, applies cohort filters, constructs redlining-zone and buffer indicators from the fields present in the CSV, executes transformations (e.g., logarithms for visualization), and produces the numeric summaries.

The **spatial layers** used to contextualize parcel membership are those prepared in §§7.1–7.2 and brought into the analysis via the CSV's precomputed flags: era-appropriate city limits (`city_limit_1945`, `city_limit_1975`, `city_limit_1985`, `city_limit_2012`, `city_limit_2020`, `city_limit_2021`) and three alternative redlining zone constructions (`RL1`, `RL2`, `RL3`) along with their ring buffers. In the workflow reported here, parcels are not re-geoprocessed anew; rather, the CSV already encodes the necessary year-specific city-status and RL membership fields, and the Python script consumes those fields as inputs.

7.3.2 Key variables and naming conventions

To maintain consistency between the manuscript and the code, the report adopts the exact field names present in the csv file and the Jupyter notebook. This subsection records the conventions so that every table and figure can be read without consulting the script.

City-limit status. Era-specific municipal membership is encoded as six indicator fields named

- `IsIn_citylimit_1945`,
- `IsIn_citylimit_1975`,
- `IsIn_citylimit_1985`,
- `IsIn_citylimit_2012`,
- `IsIn_citylimit_2020`, and
- `IsIn_citylimit_2021`.

Redlining cores and buffers. The CSV stores three alternative core definitions as `RedliningZone_1`, `RedliningZone_2`, and `RedliningZone_3`. For readability, the text refers to these as **RL1**, **RL2**, and **RL3**, respectively. For each RL variant, the dataset also includes distance-based buffer membership indicators at **0–0.5 miles** (`RLx_buf1`), **0.5–1.0 miles** (`RLx_buf2`), and **1.0–1.5 miles** (`RLx_buf3`). A fourth indicator, `RLx_buf4`, represents parcels **outside** the three rings and outside the respective RL core (i.e., located **more than 1.5 miles** from that RL variant's boundary). The analysis in Section 7 uses `buf1–buf3` to describe gradients in proximity and uses `buf4` when an explicit “Outside >1.5 mi” comparison group is required (for example, in the In-vs-Out tests in §7.6). Because the three RL variants are analyzed separately, totals reported under RL1, RL2, and RL3 are not mutually exclusive; a parcel may belong to more than one RL core across variants.

Outcome variables. Four outcome families are used throughout Section 7. For **total assessed value**, both the nominal series `TotalValue_{yr}` and the CPI-adjusted series `F2025DollarValueTotalValue_{yr}` are employed, with the inflation-adjusted series serving as the headline measure when a single metric is needed. For **per-square-foot values**, denominators are the study-year-specific fields `LandSizeSqFT_{yr}`; rows with zero lot area are excluded wherever a \$/sqft quantity is computed. When figures employ a logarithmic scale (to stabilize the visual spread of highly right-skewed distributions), non-positive observations are dropped from the visualizations, consistent with the implementation in the analysis script.

Year subscripts. The `{yr}` placeholder in field names corresponds to the six study years used throughout this section: **1945**, **1975**, **1985**, **2012**, **2020**, and **2021**. When a statement applies to all years, the notation `_ {yr}` is used to denote the family of fields in the CSV.

7.3.3 Cohort construction

Section 7 reports statistics under three complementary cohort constructions so that patterns observed in per-year snapshots can be compared with patterns observed on balanced panels.

- The **Baseline cohort** serves as the primary sample. It is constructed separately for each study year and includes all parcels present in the CSV for that year, with city membership taken from the corresponding `IsIn_citylimit_{yr}` field. This cohort maximizes coverage and is therefore used for most descriptive summaries (counts, medians, interquartile ranges, and boxplots) within a given year. Because this cohort is re-defined for each year, membership can and does change across years as annexations and data availability change. When the text refers to “Baseline counts,” it refers to the per-year universe that satisfies the year’s inclusion conditions as encoded in the CSV and consumed by the script.
- The **Longitudinal6 cohort** is a balanced panel across **all six** study years. Parcels are included in Longitudinal6 if the fields required for the specified outcomes are **non-null in every year** and satisfy the additional conditions enforced by the script (for example, the exclusion of zero lot area where \$/sqft metrics are computed, and any other filters embedded in the workflow). Longitudinal6 supports direct within-parcel comparisons over the full 1945–2021 span. Because this cohort trades breadth for strict temporal completeness, its counts are necessarily smaller than Baseline counts in any single year.
- The **Longitudinal5 cohort** is a balanced panel across the **latest five** study years—1975, 1985, 2012, 2020, and 2021—constructed with the same logic as Longitudinal6 but without the 1945 requirement. This cohort improves sample size over Longitudinal6 while still enabling long-horizon comparisons that bridge the archival and contemporary eras covered by the dataset. As in Longitudinal6, the manuscript adheres to the script’s filtering logic when determining inclusion.

Two clarifications apply to all three cohorts. First, because RL1, RL2, and RL3 are treated as **alternative** definitions evaluated in separate analytical passes, a parcel may appear in multiple RL-specific outputs when those outputs are viewed side-by-side; this does not imply duplication within a single analysis. Second, when the analysis switches from totals to per-square-foot outcomes, rows with zero `LandSizeSqFT_{yr}` are dropped for the affected year(s), which can lead to small differences in cohort counts between outcome families; these differences reflect the dataset and the script’s rules rather than additional filtering by the manuscript.

7.3.4 Transformations and exclusions

Because assessed values are highly right-skewed, the script renders time-series visualizations on a logarithmic scale for interpretability. In keeping with standard practice and to match the code’s behavior, non-positive observations are omitted from log-based graphics; this affects plots only and does not alter the underlying tabulations for medians and interquartile ranges unless explicitly noted in a figure caption. For clarity of presentation, boxplots are produced without outlier fliers so that central tendency and dispersion within each group remain legible across years and RL variants. No winsorization or trimming of extreme values is performed in the summary statistics unless a specific figure calls for it and the code implements it; in such cases the caption will restate the exact rule used by the script.

Cohort membership interacts with these transformations in straightforward ways. Baseline cohorts are computed separately for each study year and therefore reflect the year-specific availability of outcomes and city-limit flags. Longitudinal6 and Longitudinal5 cohorts require non-null outcomes in all of their respective years; when per-sqft outcomes are reported, rows with zero lot area are dropped in the affected years, which can lead to minor deviations in counts between total-value panels and

per-sqft panels. Each table and figure in later subsections indicates which cohort and which outcome family it uses so that exclusions attributable to logs or zero-area filtering are transparent.

7.3.5 Statistical testing and notation

Within each study year and RL variant, the distribution of an outcome inside the RL **core** (RLx) is compared to the distribution **outside** that RL variant. The test used is the Mann–Whitney U test, a nonparametric procedure appropriate for comparing central tendencies between two independent samples when distributional normality and equal variances cannot be assumed. Tests are run separately by year, by cohort (Baseline, Longitudinal6, Longitudinal5), by RL variant (RL1, RL2, RL3), and by outcome family (nominal totals, inflation-adjusted totals, and their per-sqft counterparts).

To keep figures readable while preserving the underlying numeric results, the manuscript adopts the same compact p-value notation. Each line chart that juxtaposes **Core** and **Outside** medians or means is annotated above the year tick with a glyph that summarizes the Mann–Whitney U result for that year. The glyphs used are **VS (very significant)**, **S (significant)**, **B (border)**, and **VI (very insignificant)**, representing four p-value bands as defined in the code. Rather than redefining thresholds here, the legend will be inserted verbatim from the script's mapping so that the figure captions and the glyph meanings remain exactly aligned with the project's implementation.

IN vs OUT Time-Series by Cohort and Statistic (RL1–RL3; Four Outcomes)

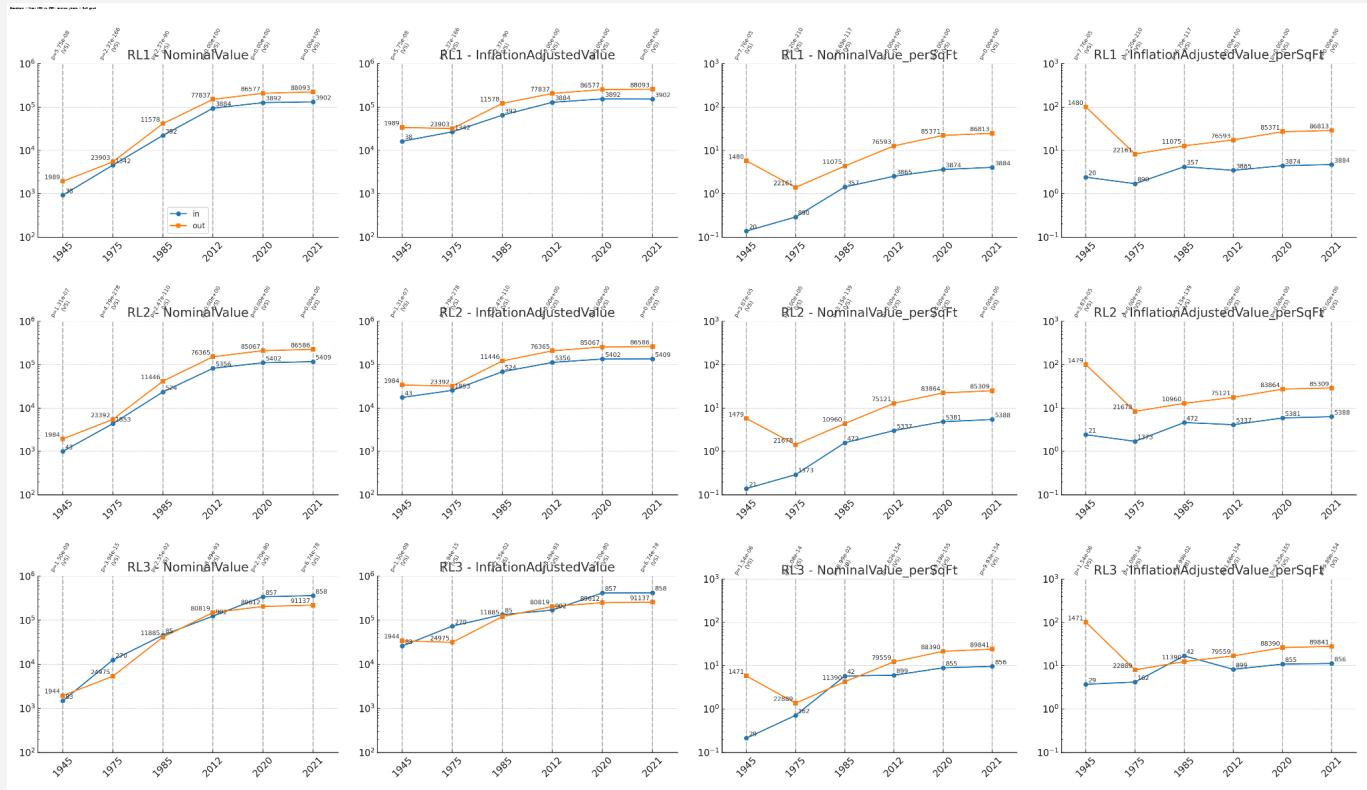


Figure 7.5 — Baseline — Mean (in vs out) RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.

Across all four outcomes, **OUT** sits above **IN** for every RL. For **Nominal** and **Inflation-adjusted totals**, **RL2** edges **RL1** by a small average OUT/IN ratio (≈ 1.77 vs ≈ 1.68), but the lines are close and often marked **S/VS**. For **per-sqft** outcomes, **RL1** shows the **largest separation** (average OUT/IN ≈ 11.1 vs ≈ 10.5 for RL2), indicating denser valuation penalties inside RL1 when normalized by parcel

area. **RL3** consistently shows the smallest gaps. Interpretation: on mean levels, RL2 is a very close competitor on totals, but **RL1 is the stronger and more consistent “most affected” footprint once normalized by size.**

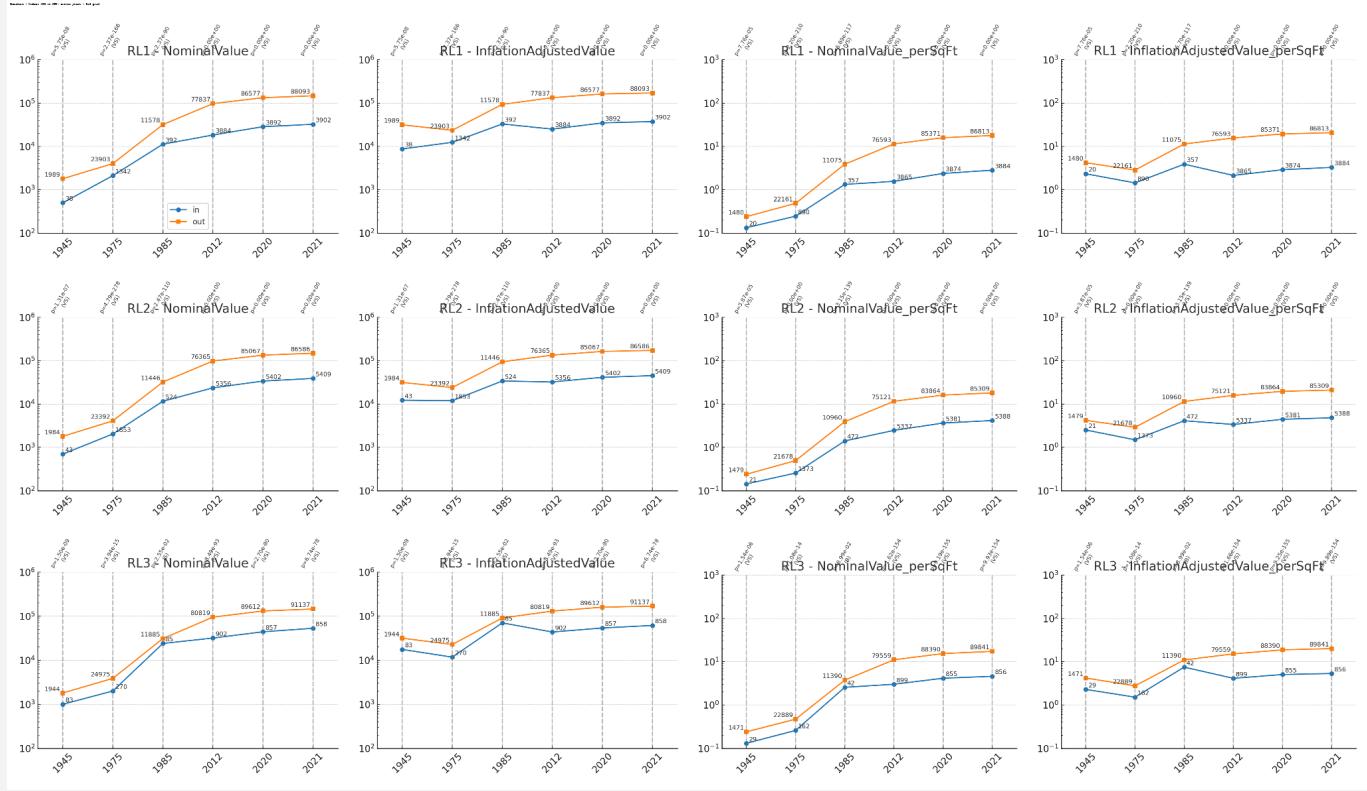


Figure 7.6 — Baseline — Median (in vs out) RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.

Medians accentuate central-tendency differences while damping outlier influence. Here **RL1 clearly dominates**: average OUT/IN ≈3.81 for both nominal and inflation-adjusted medians, exceeding RL2 (≈3.22) and RL3 (≈2.29). For **per-sqft medians**, RL1 again leads (≈4.51 vs ≈3.31 for RL2). P-value markers are predominantly **S/VS** in late years. Interpretation: on medians (the main narrative for skewed value data), **RL1 is the most affected variant** in Baseline.

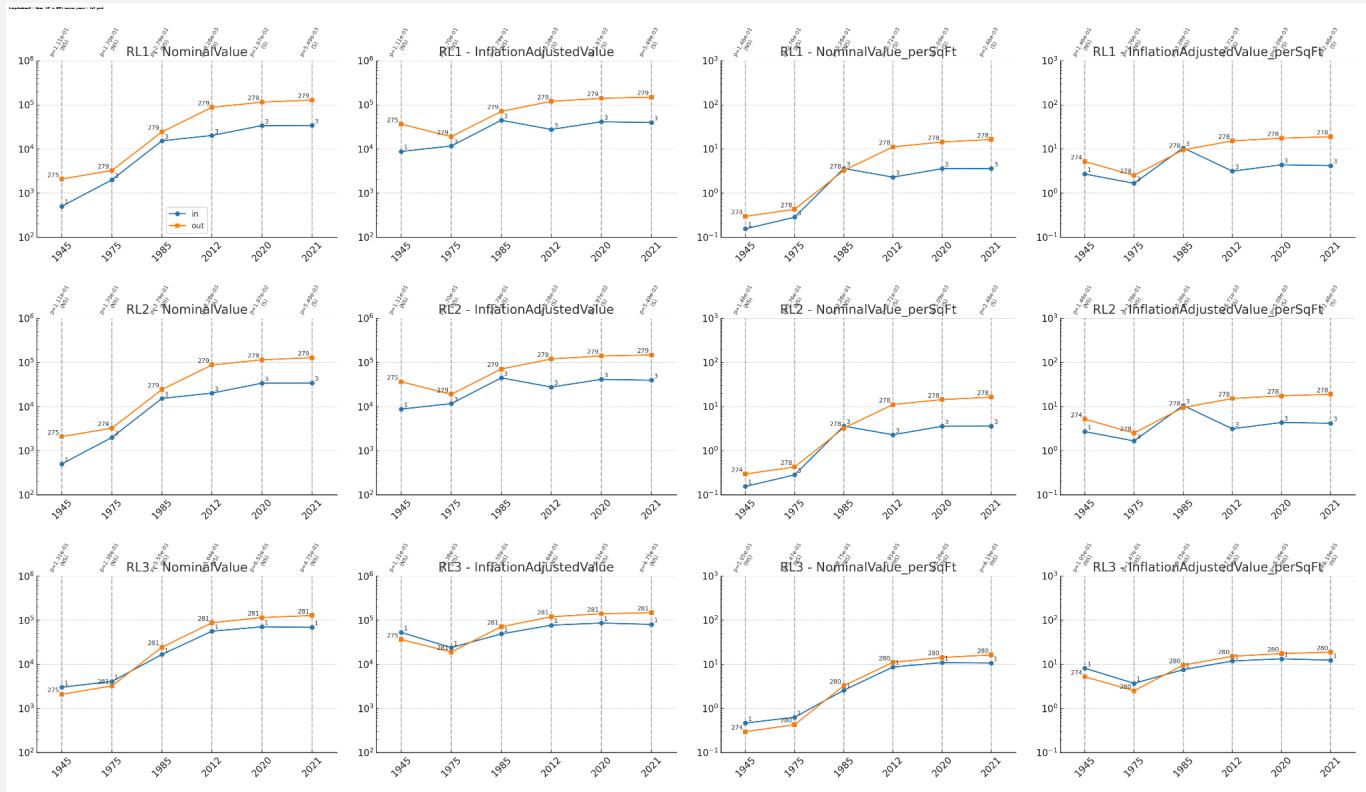


Figure 7.7 — Longitudinal6 — Mean (in vs out) Six-year balanced panel; RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.

Using the strict six-year balanced panel, **RL1** has the **largest average OUT/IN** for both totals (≈ 3.15) and per-sqft (≈ 2.96), with RL2 slightly behind and RL3 well below 2x. Many years retain S/VS markers despite reduced n, indicating the disparity is not an artifact of changing parcel composition. Interpretation: even under the most conservative cohort, **RL1 remains the most affected.**

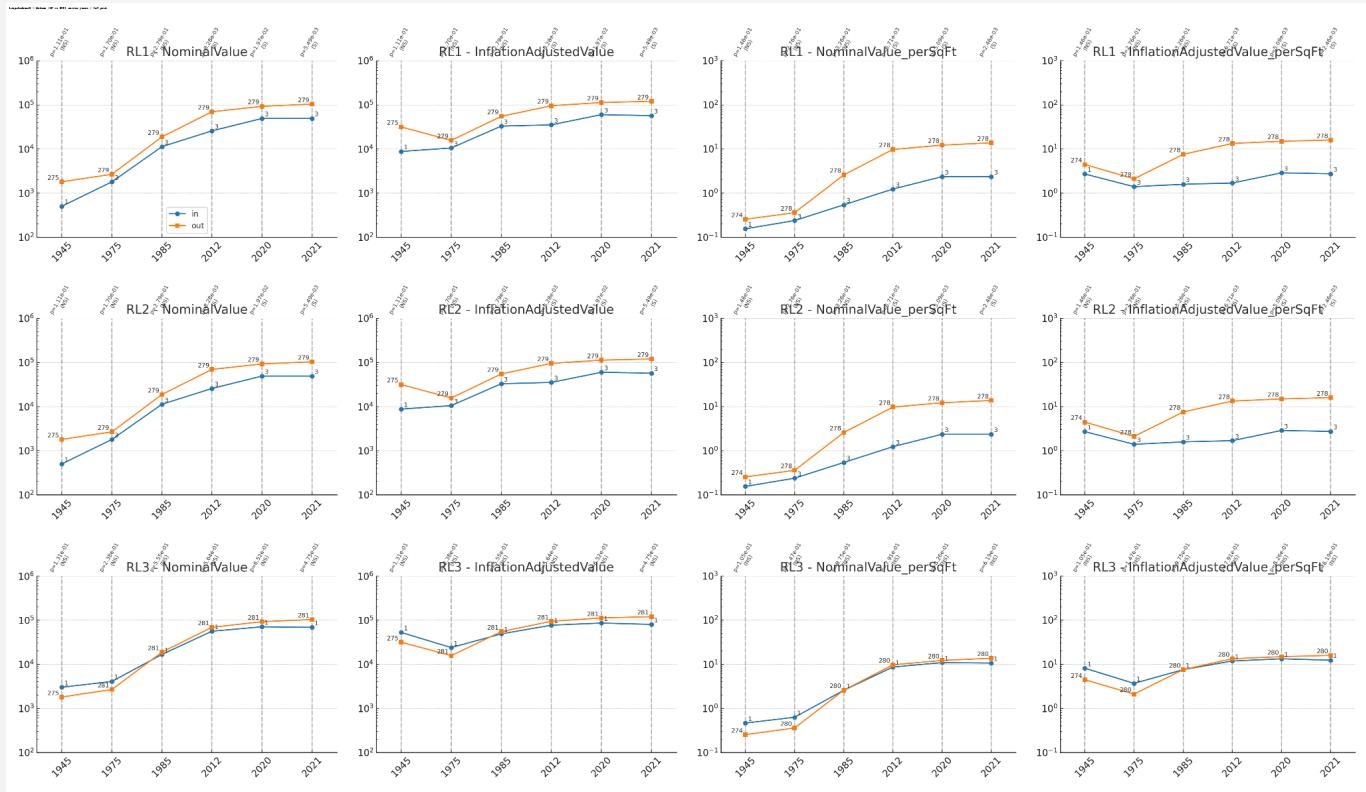
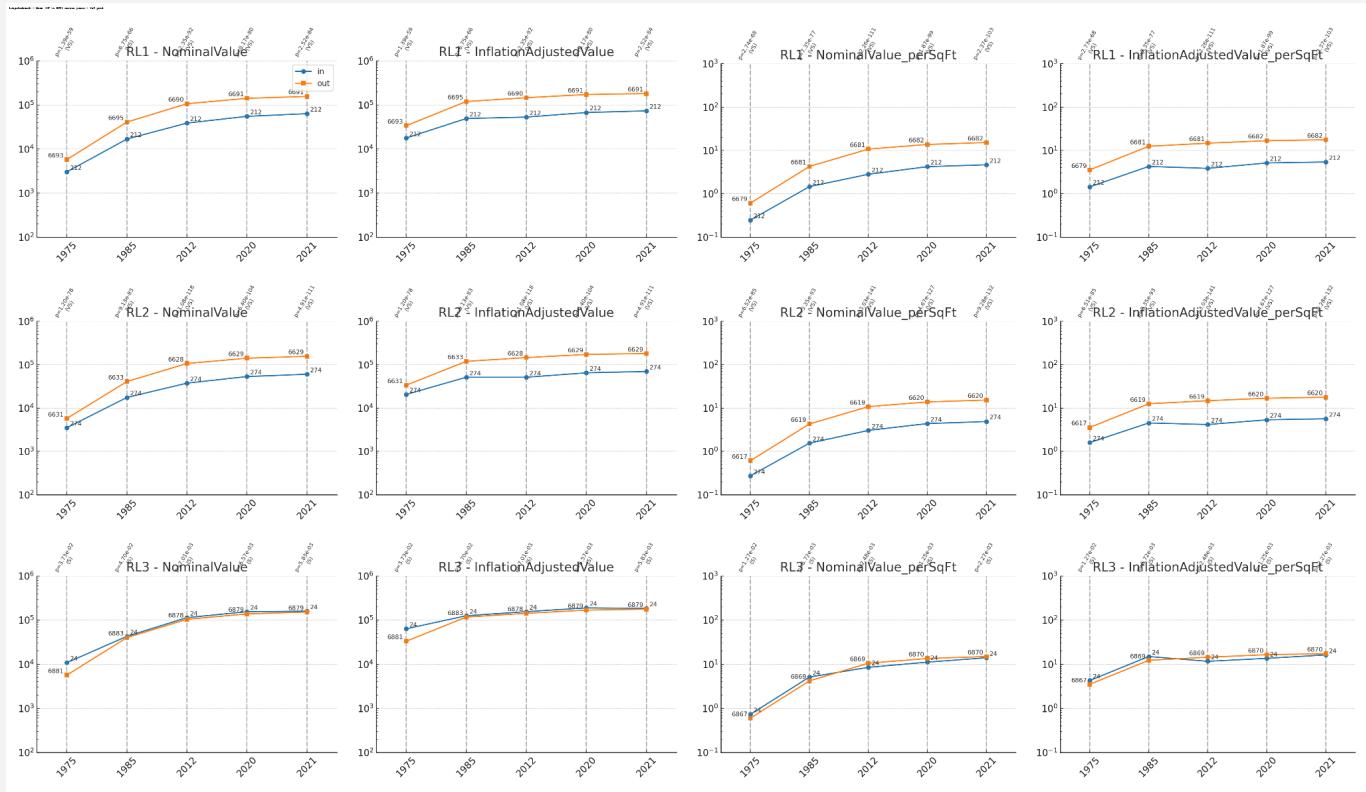


Figure 7.8 — Longitudinal6 — Median (in vs out) Six-year balanced panel; RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.

For medians on the six-year panel, **RL1** again leads across totals (≈ 2.24) and per-sqft (≈ 4.48), with **RL2** second and **RL3** the smallest. Significance markers frequently show **S/V/S**. Interpretation: **RL1** provides the largest core penalty at the distribution's center once panel consistency is enforced.



**Figure 7.9 — Longitudinal5 — Mean (in vs out) 1975, 1985, 2012, 2020, 2021; RL1–RL3;
Nominal, Inflation-Adjusted, \$/sqft.**

On the five-year panel (1975+), **RL1** has the highest mean ratios for both totals (≈ 2.42 , slightly above $RL2 \approx 2.41$) and per-sqft (≈ 3.14 vs $RL2 \approx 2.97$). The differences vs $RL2$ are modest but persistent across years, with many S/VS p-values. Interpretation: **RL1** edges $RL2$ on means once attention is restricted to the recent decades.

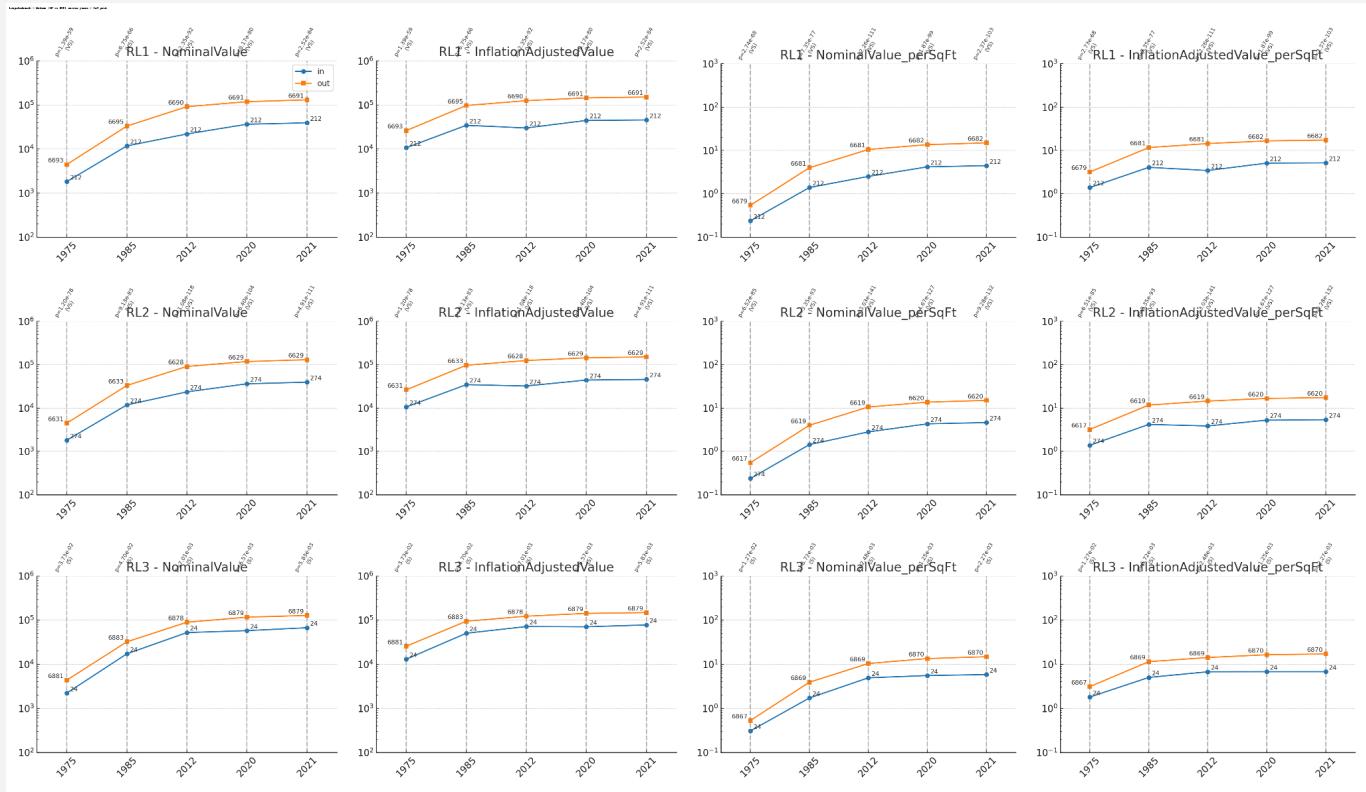


Figure 7.10 — Longitudinal5 — Median (in vs out) 1975, 1985, 2012, 2020, 2021; RL1–RL3; Nominal, Inflation-Adjusted, \$/sqft.

Median contrasts are strongest and most decisive: **RL1** posts the largest OUT/IN for totals (≈ 3.20) and per-sqft (≈ 3.19), with **RL2** clearly lower and **RL3** the smallest. P-value stripes frequently read **S/VS**. Interpretation: for the recent-decades panel, **RL1** is unambiguously the **most affected** definition at the center of the distribution.

Bottom line Summary

Across cohorts (Baseline, Longitudinal6, Longitudinal5), statistics (mean, median), and outcomes (nominal, inflation-adjusted, and per-sqft), the **Inside vs Outside** time-series grids show a stable hierarchy: **RL1** exhibits the **largest and most consistently significant** disparity, **RL2** is a close but generally smaller second, and **RL3** is the smallest. The preference for **RL1 as the operative redlining footprint** strengthens when focusing on **medians** and **per-sqft** outcomes (which are less sensitive to outliers and lot size), and it persists under the strict six-year balanced panel—evidence that the disparity is not driven by shifting parcel membership.

7.3.6 Box Plots of three cohorts

7.3.6.1 RLs vs Cohorts

Layout per mosaic:

- **Rows (top→bottom):** Baseline, Longitudinal6, Longitudinal5
- **Columns (left→right):** RL1, RL2, RL3
- **Each panel:** two box charts (no fliers) comparing **Inside RL (“in”)** vs **Outside RL (“out”)** for that zone and cohort in the specified year. The y-axis uses **symlog** scaling. Counts and summary stats (max, q75, mean, median, q25, min) are annotated on each box.

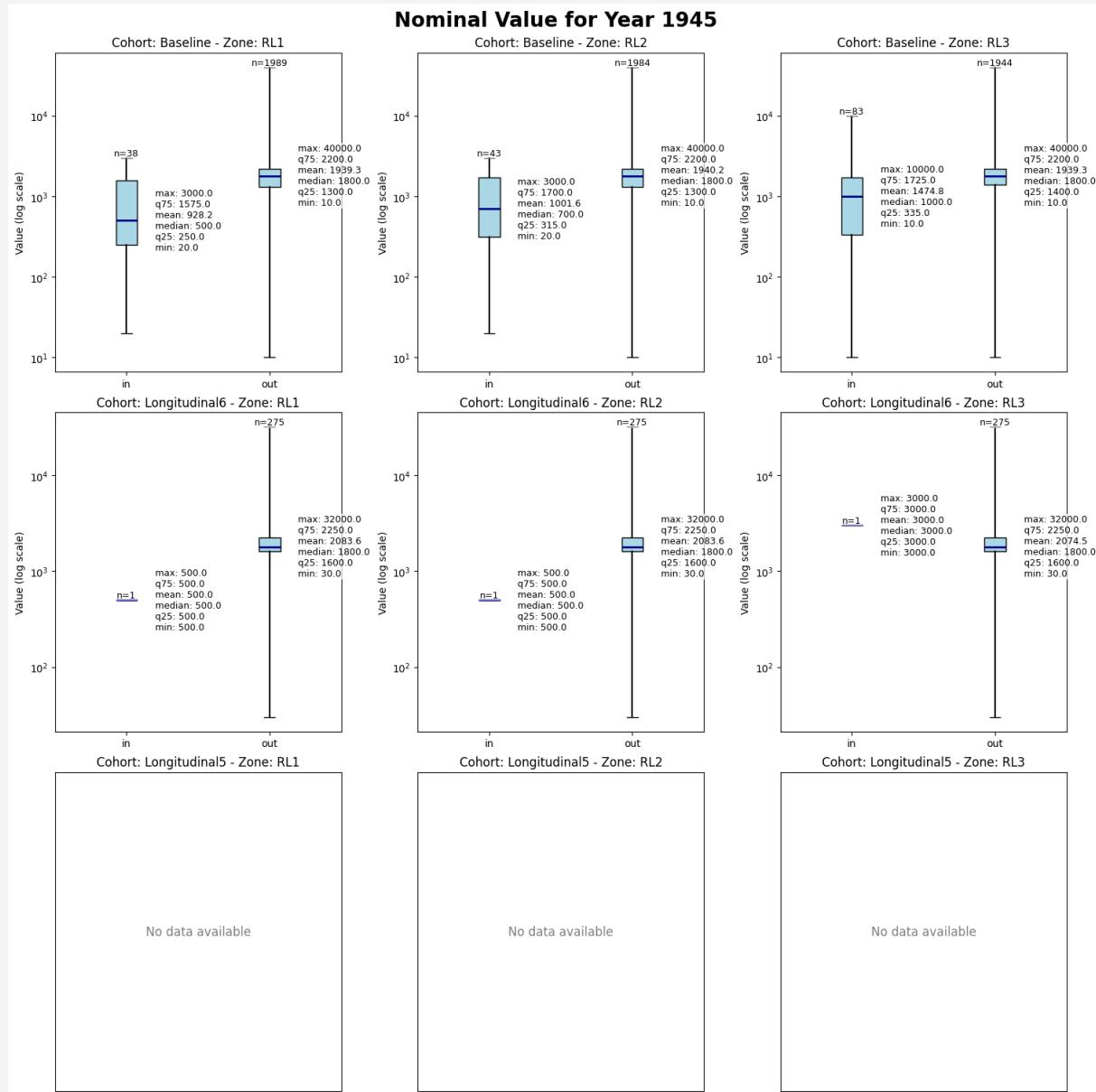


Figure 7.11 — Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 1945

Baseline: Median **inside** the RL cores is markedly below **outside** for RL1 (≈ 500 vs $1,800$) and RL2 (≈ 700 vs $1,800$); RL3 shows $\approx 1,000$ vs $1,800$. Counts are small inside cores relative to outside,

especially for RL1/RL2 (e.g., **Count_IN** dozens vs **Count_OUT** ~1,900+), consistent with a concentrated core footprint in this early year.

Longitudinal6: Extremely small **in** counts (often **n = 1**), producing step-like medians; in RL3 the single inside observation yields a higher “in” median ($\approx 3,000$) than “out” ($\approx 1,800$), which should be interpreted cautiously given **n = 1**.

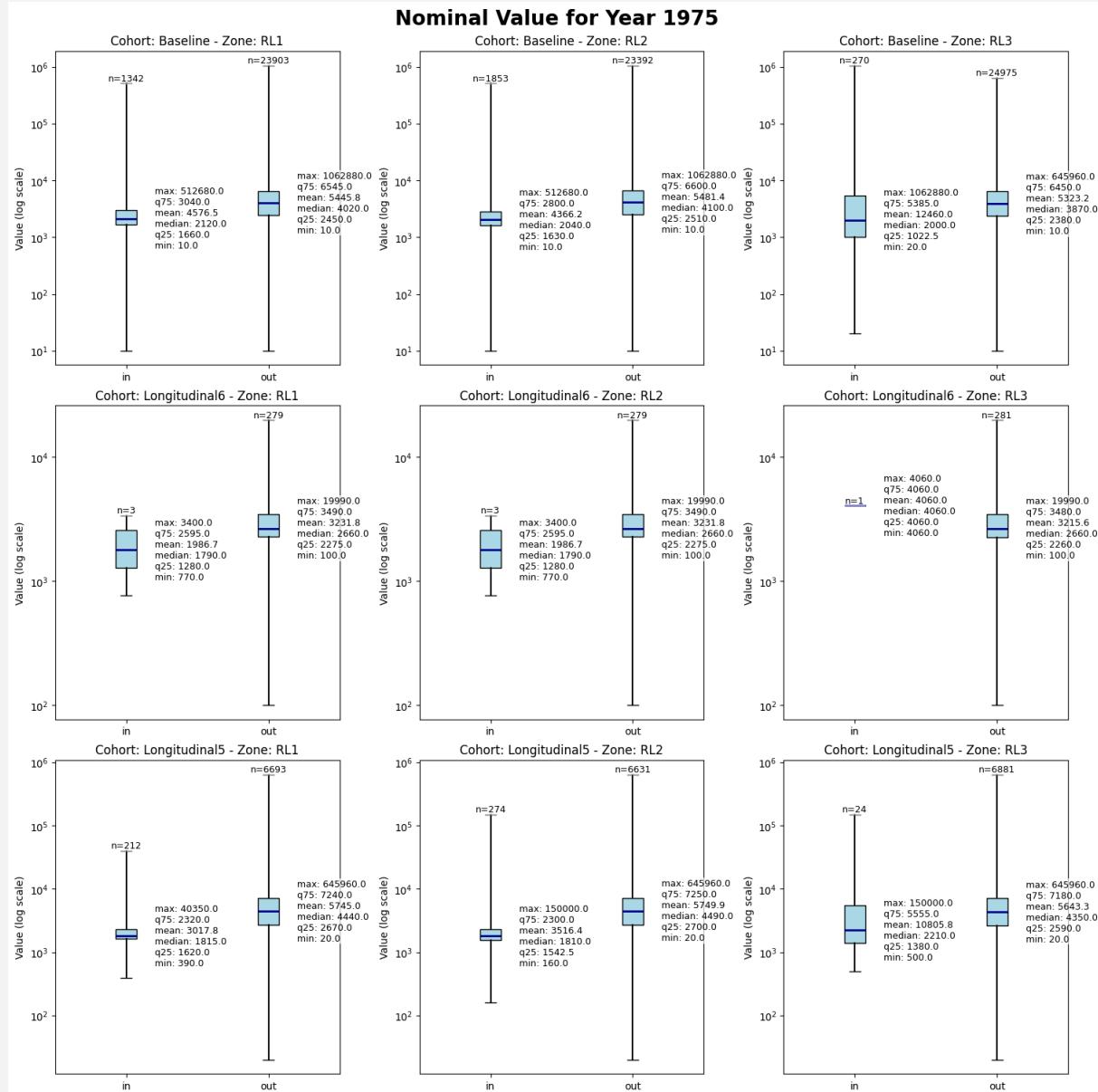


Figure 7.12 — Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 1975

Baseline: A stable **inside < outside** gap across all RL variants—medians near **2.0–2.1k** for **in** vs **~4.0–4.1k** for **out** (RL1: $\approx 2,120$ vs $4,020$; RL2: $\approx 2,040$ vs $4,100$; RL3: $\approx 2,000$ vs $3,870$). Inside counts are now in the high hundreds to low thousands for RL1/RL2, supporting robust medians; RL3 remains smaller.

Longitudinal6: Medians rise for both groups but preserve the same ordering (e.g., RL1 **in** $\approx 1,790$ vs **out** $\approx 2,660$).

Longitudinal5: Gaps resemble Baseline (e.g., RL1 **in** \approx 1,815 vs **out** \approx 4,440; RL2 **in** \approx 1,810 vs **out** \approx 4,490; RL3 **in** \approx 2,210 vs **out** \approx 4,350), now with larger **in** counts that stabilize the summaries.

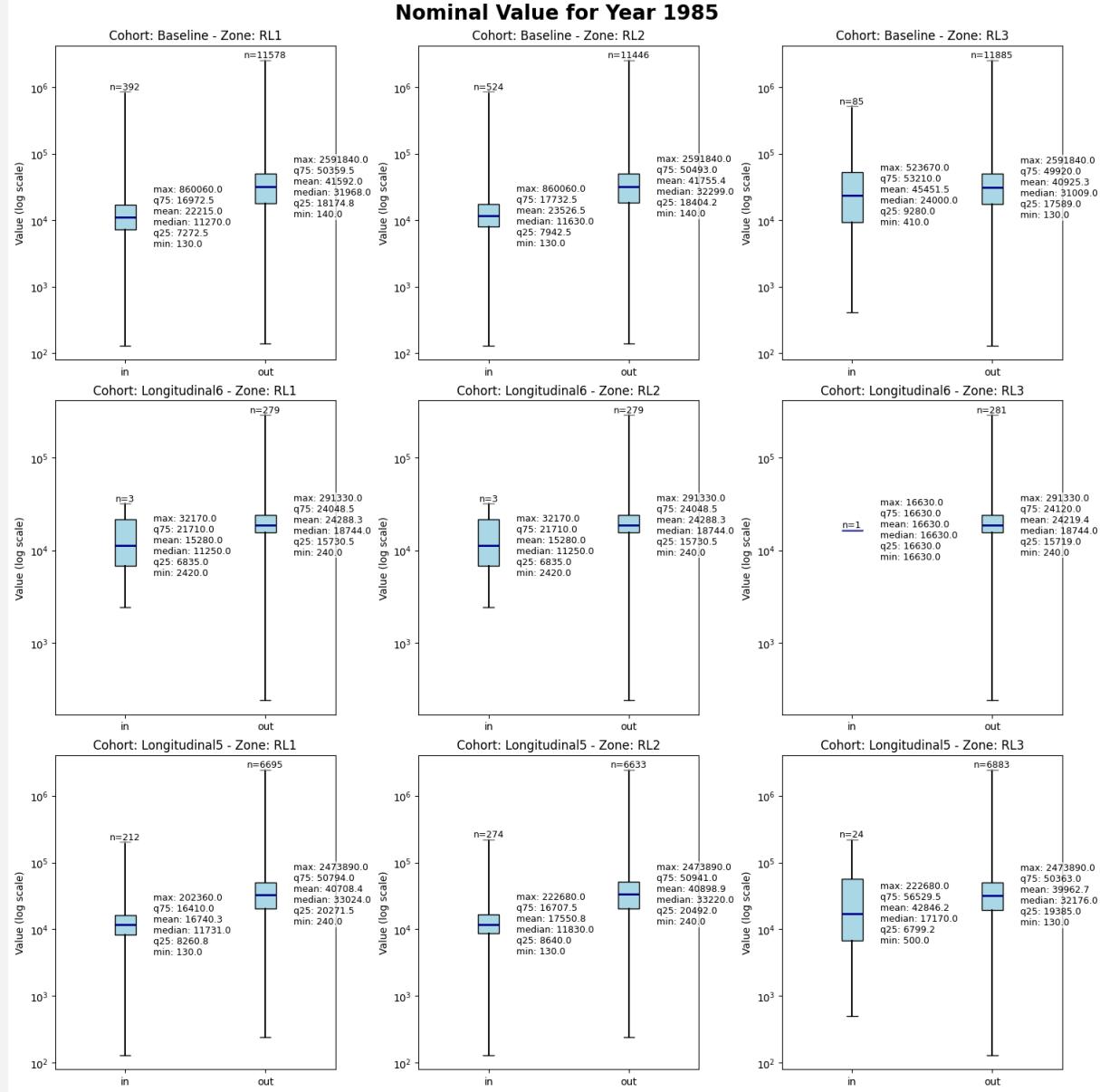


Figure 7.13 — Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 1985

Baseline: Level shift upward across city and zones. Medians remain **inside < outside**: RL1 \approx 11.3k vs 31.97k; RL2 \approx 11.6k vs 32.30k; RL3 \approx 24.0k vs 31.01k. RL3 inside medians are higher than RL1/RL2 but still trail the outside group.

Longitudinal6: Same ordering with smaller medians than Baseline (e.g., RL1 \approx 11.25k vs 18.74k), reflecting the stricter balanced-panel filter.

Longitudinal5: Intermediate magnitudes and stable gaps (e.g., RL1 \approx 11.73k vs 33.02k; RL2 \approx 11.83k vs 33.22k; RL3 \approx 17.17k vs 32.18k).

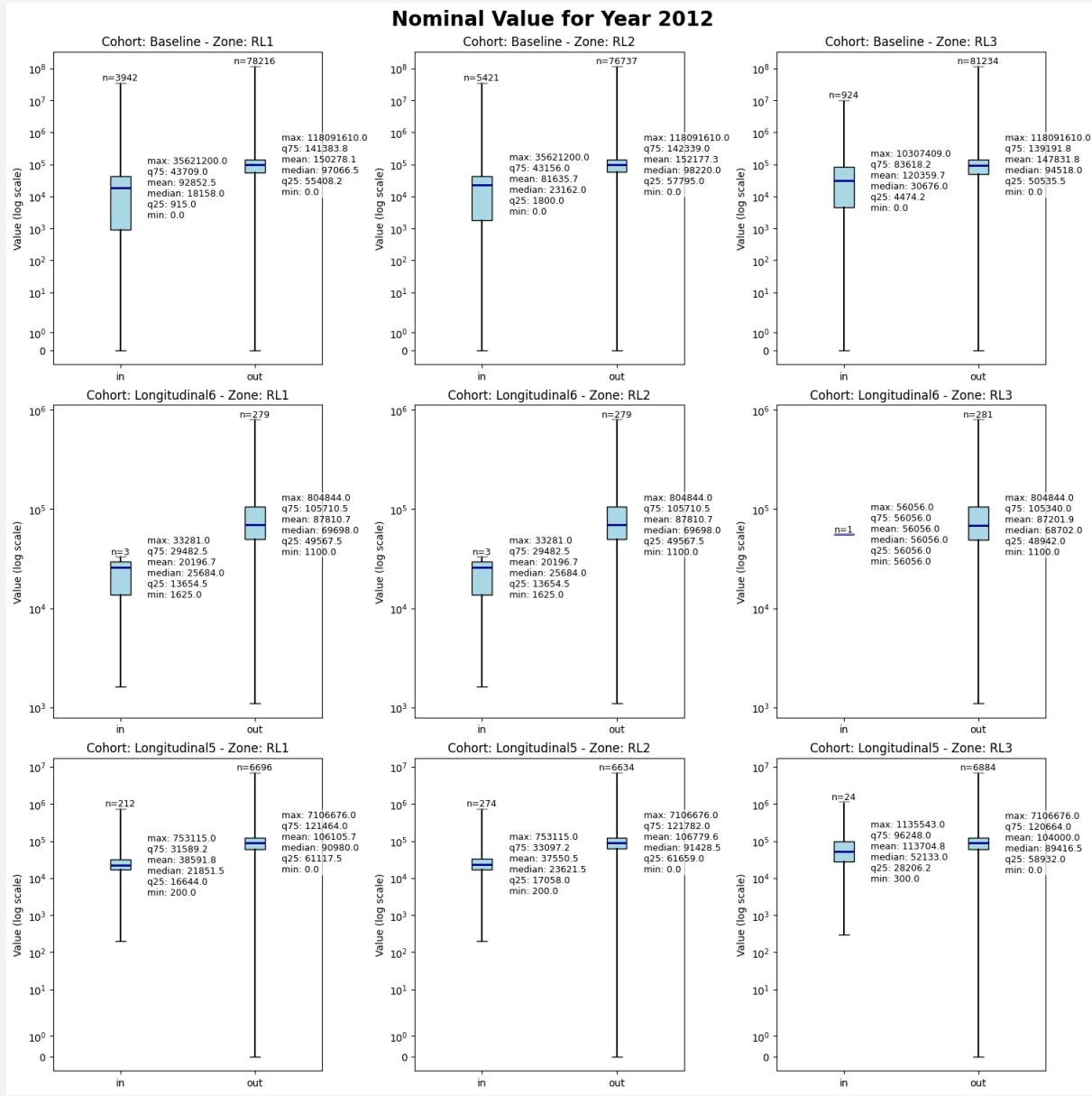


Figure 7.14 — Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 2012

Baseline: Medians increase steeply citywide. **Inside** remains far below **outside**: RL1 \approx 18.16k vs 97.07k; RL2 \approx 23.16k vs 98.22k; RL3 \approx 30.68k vs 94.52k. Inside counts are now in the thousands for RL1/RL2 and near a thousand for RL3.

Longitudinal6: The same pattern with lower absolute levels (e.g., RL1 \approx 25.68k vs 69.70k; RL3 \approx 56.06k vs 68.70k), tracking the smaller, consistent sample.

Longitudinal5: Patterns mirror Baseline (e.g., RL1 \approx 21.85k vs 90.98k; RL2 \approx 23.62k vs 91.43k; RL3 \approx 52.13k vs 89.42k).

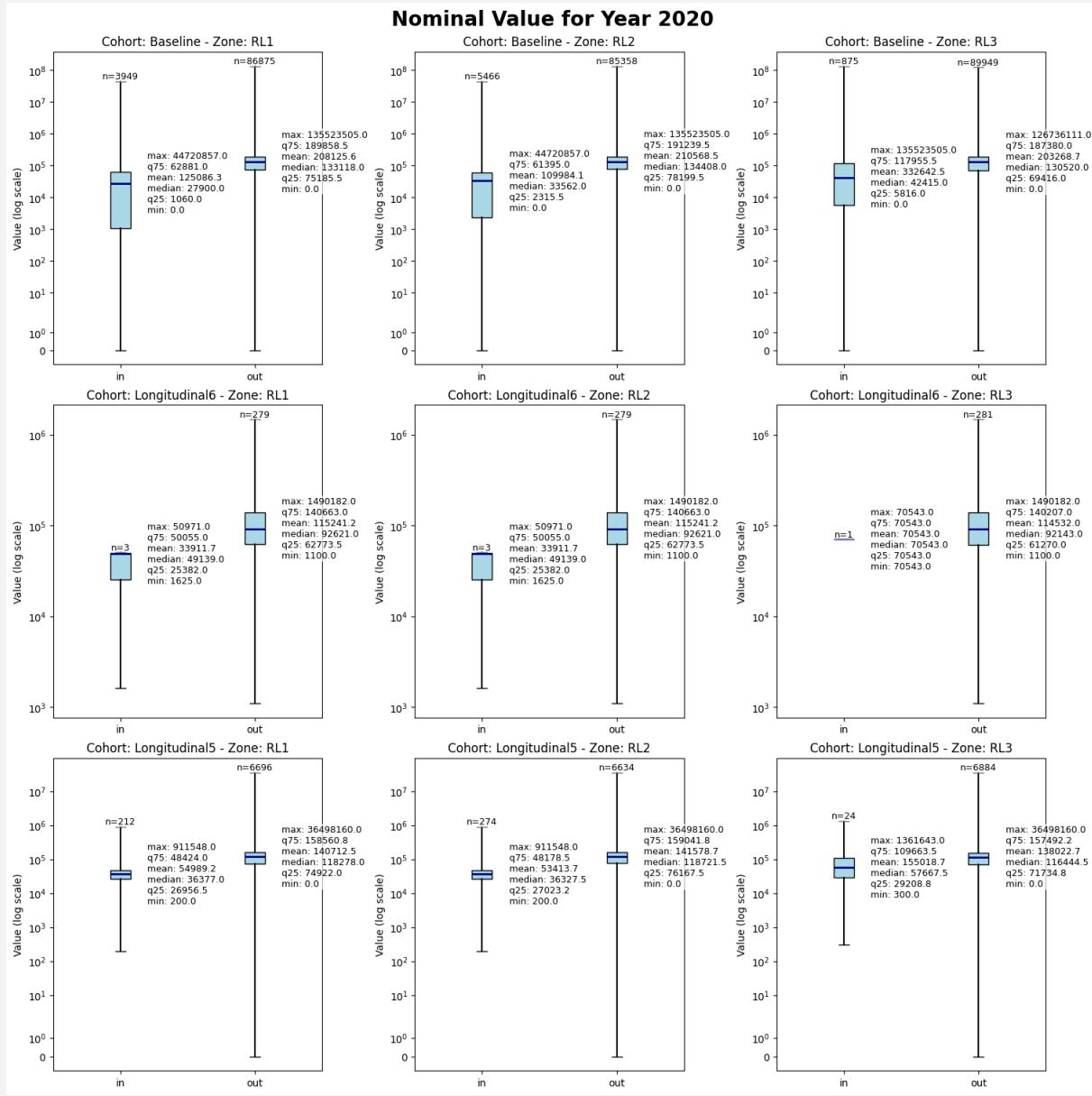


Figure 7.15 — Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 2020

Baseline: Very large separation remains: RL1 \approx 27.90k vs 133.12k; RL2 \approx 33.56k vs 134.41k; RL3 \approx 39.50k vs 130.52k. The outside group dominates counts and levels.

Longitudinal6 / Longitudinal5: The gap persists under balanced panels (e.g., RL1 Long5 \approx 36.38k vs 118.28k; RL3 Long5 \approx 57.67k vs 116.44k), supporting the conclusion that the disparity is not solely composition-driven.

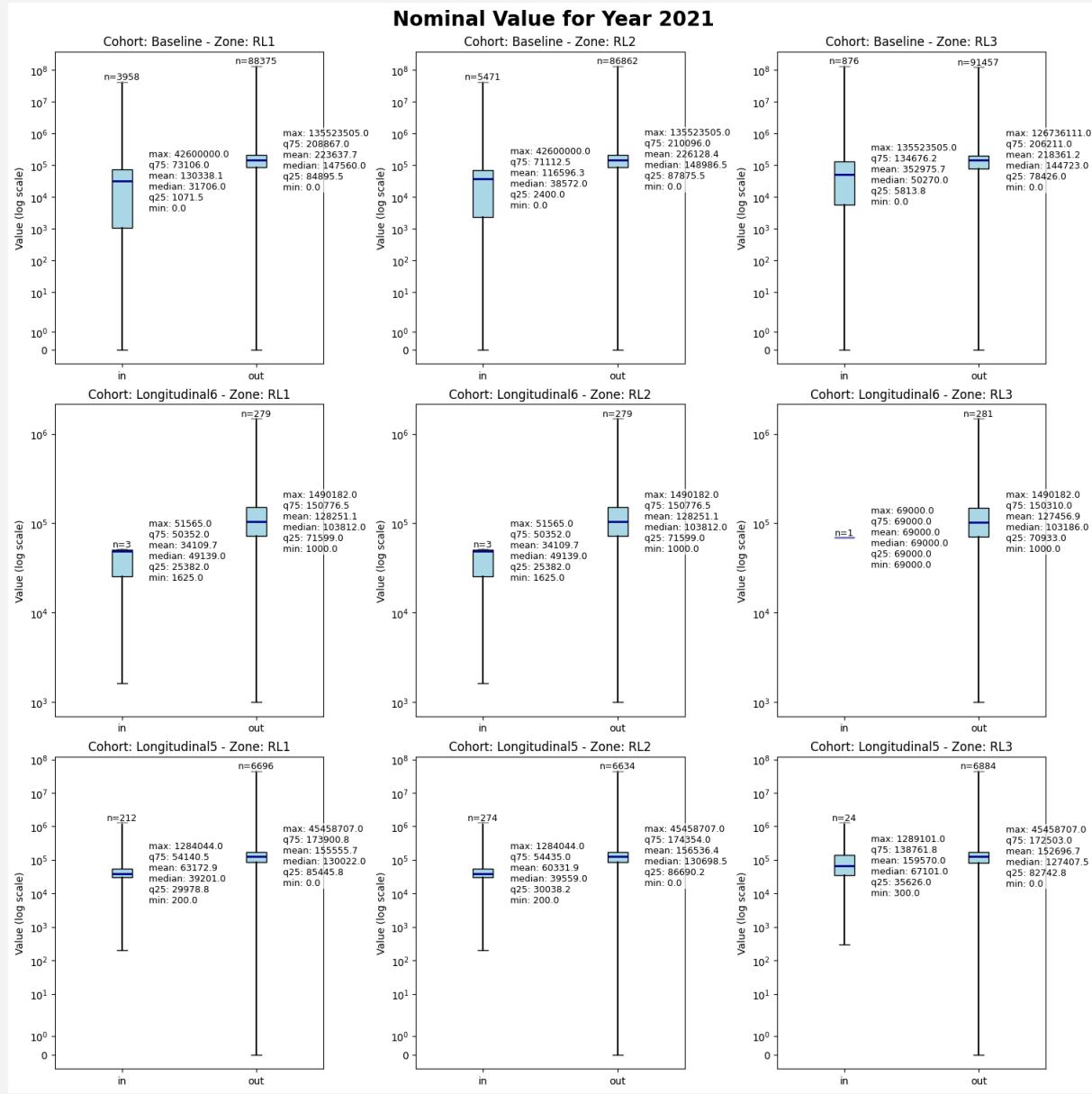


Figure 7.16 — Nominal total value — Inside vs Outside RL (rows: Baseline, Longitudinal6, Longitudinal5; columns: RL1, RL2, RL3), Year 2021

Baseline: The largest observed medians citywide; **inside** remains substantially below **outside** across all RL variants (RL1 \approx 30.55k vs 147.56k; RL2 \approx 35.80k vs 148.99k; RL3 \approx 41.06k vs 144.72k).

Longitudinal6 / Longitudinal5: The same ordering holds (e.g., RL1 Long5 \approx 39.20k vs 130.02k; RL3 Long5 \approx 67.10k vs 127.41k), reinforcing persistence into the most recent year.

7.3.6.2 Groups (in vs out) vs RLs

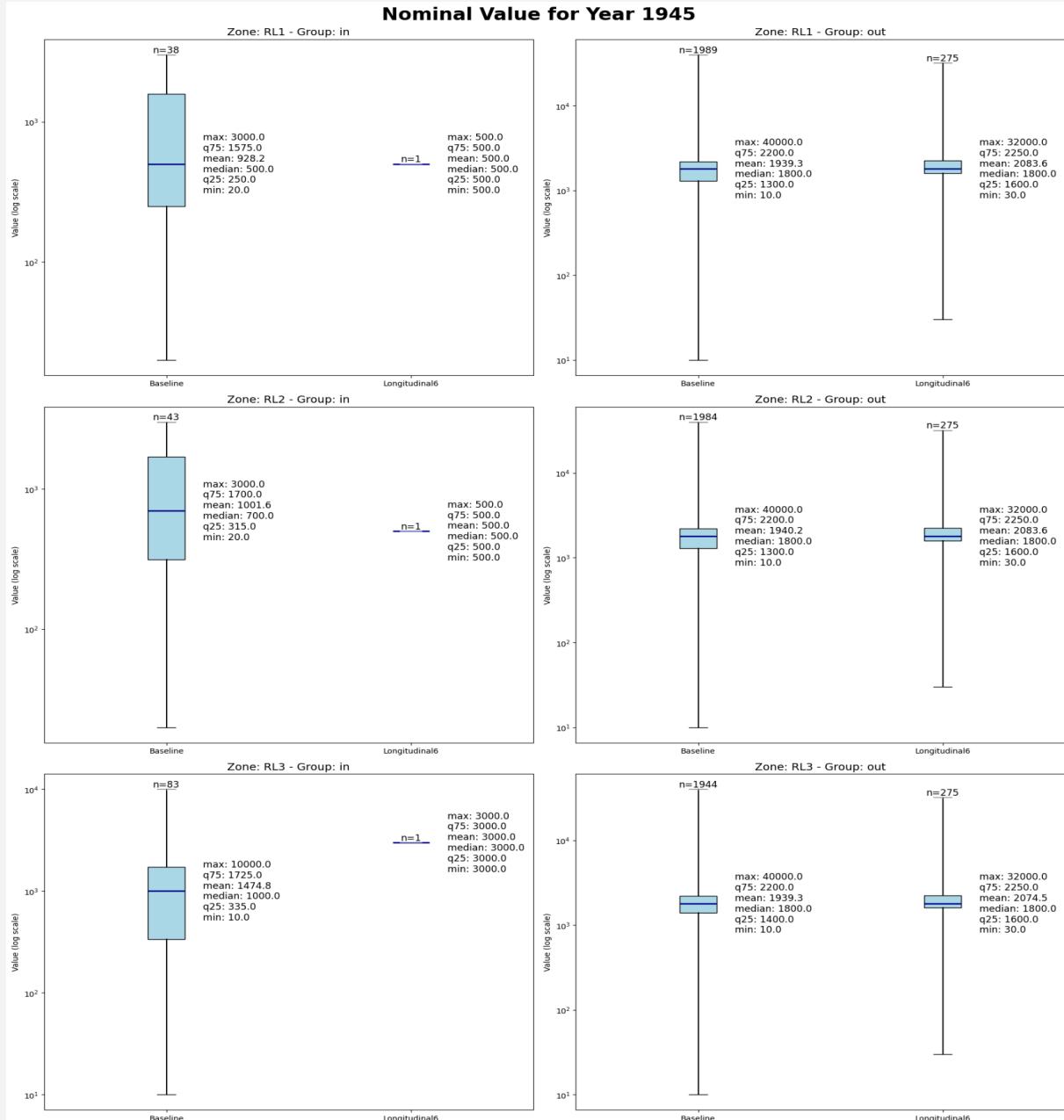


Figure 7.17 — Nominal Value (TotalValue) — Year 1945 — cohorts by zone \times group

Inside-vs-outside gaps are already visible in 1945: across RL1 and RL2, inside medians sit well below outside medians; RL3 shows a very small inside cell with correspondingly volatile summaries. Baseline boxes dominate counts, while the Longitudinal6 row presents much smaller n's and step-like medians consistent with a strict balanced-panel filter. As an early baseline, the figure indicates that lower assessed levels inside the candidate redlining cores were present at the start of the study horizon. Given small inside-cell counts (especially for RL3), emphasis should remain on direction rather than precise magnitudes for this year.

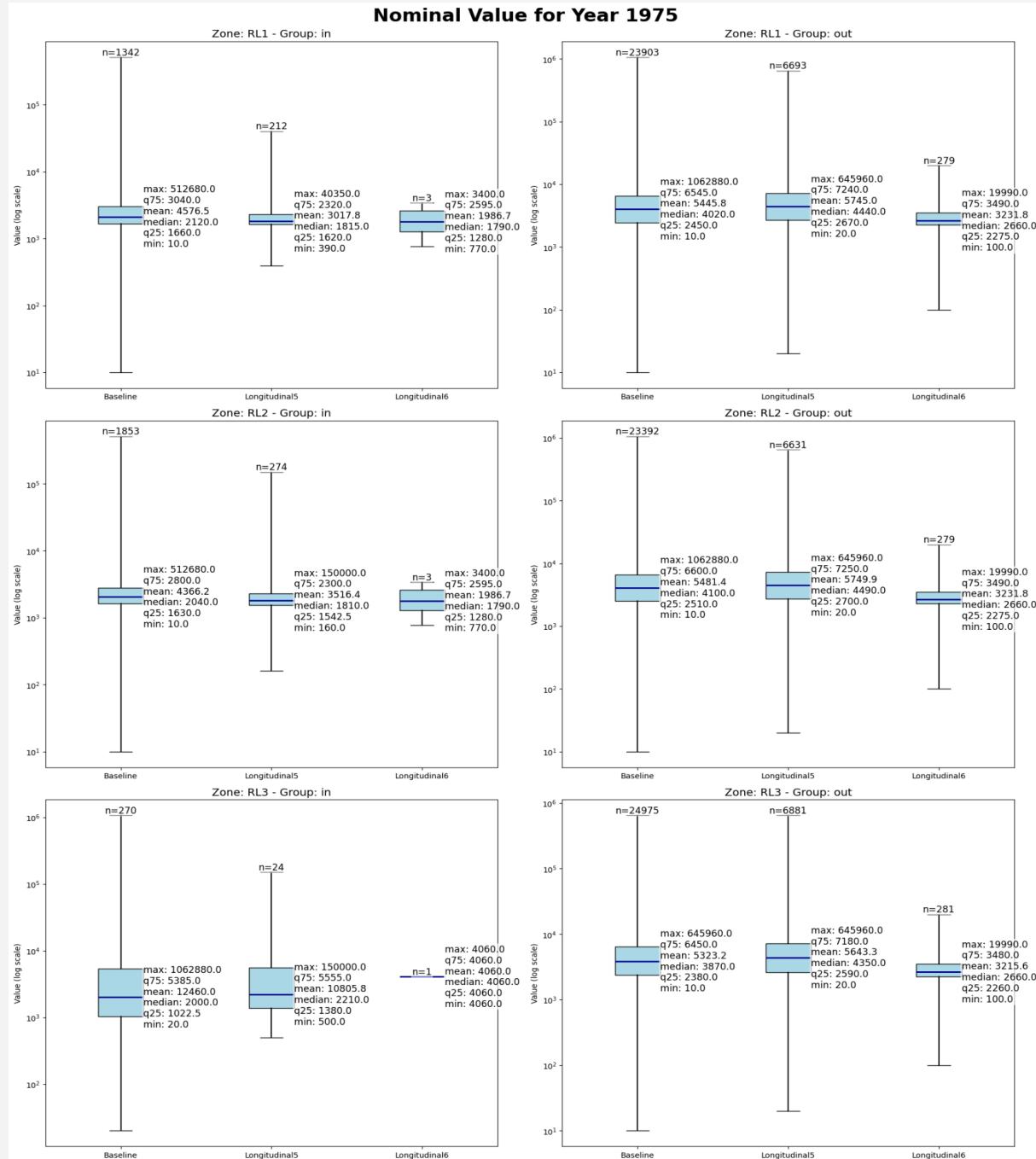


Figure 7.18 — Nominal Value (TotalValue) — Year 1975 — cohorts by zone \times group

By 1975, the inside < outside ordering remains clear for all RL variants. Relative to 1945, counts inside cores are larger for RL1 and RL2, stabilizing medians and interquartile ranges; RL3 remains compact. Longitudinal5 boxes (panel row 3) closely track Baseline shapes, suggesting that the observed gaps are not purely compositional and persist within a balanced late-period panel. Longitudinal6 shows the same direction with tighter spreads and slightly lower central values, consistent with stricter inclusion.

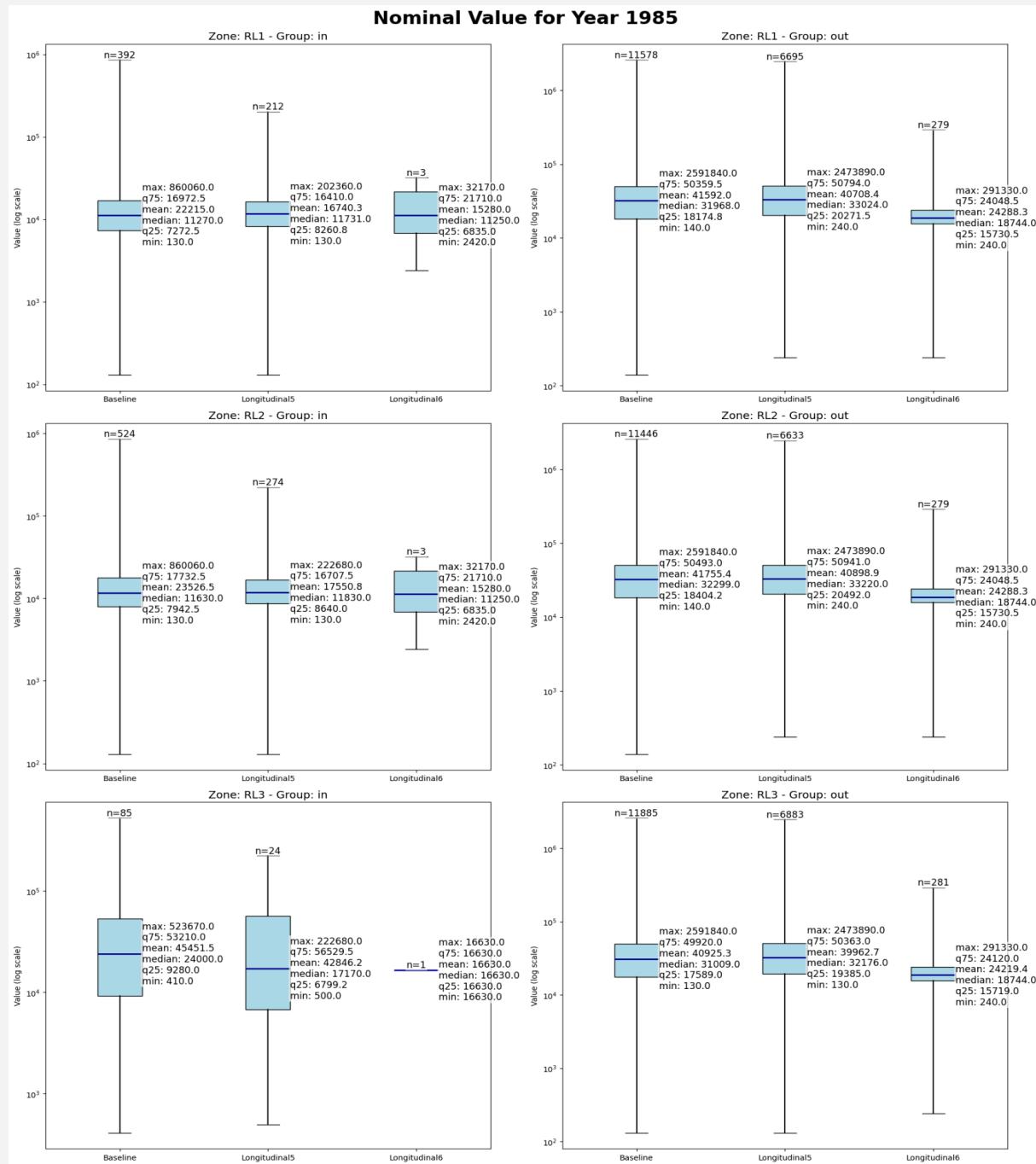


Figure 7.19 — Nominal Value (TotalValue) — Year 1985 — cohorts by zone×group

Citywide levels rose sharply by 1985 (nominal terms), yet the **inside < outside** pattern persists under RL1, RL2, and RL3. The RL3 “inside” box tends to sit above RL1/RL2 “inside” but still below the corresponding “outside,” illustrating sensitivity to the RL definition without changing the direction of the disparity. Longitudinal5 again mirrors Baseline contrasts, while Longitudinal6 remains more compact with smaller n’s. Overall, the figure supports continuity of the within-year gap through mid-1980s conditions.

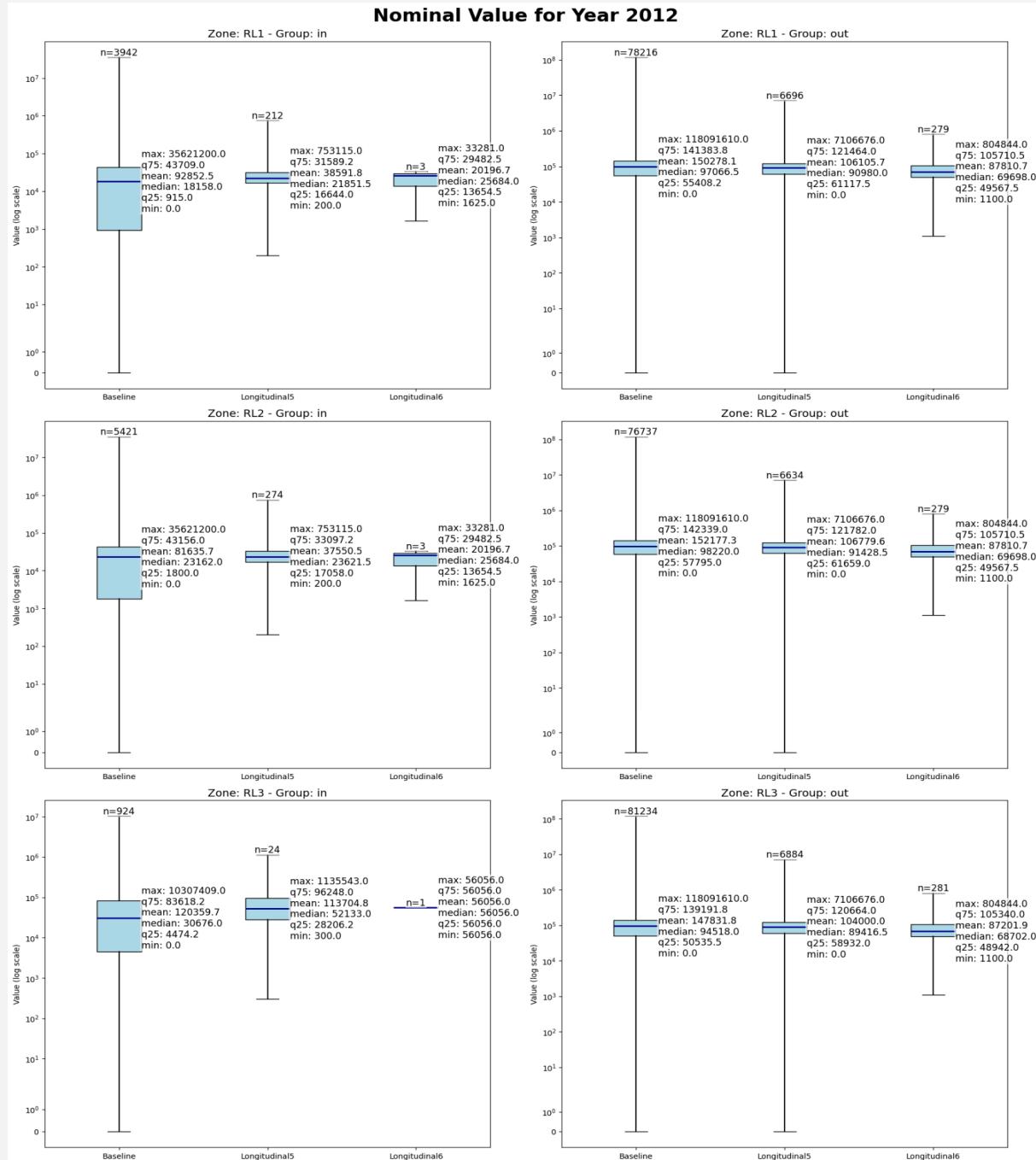


Figure 7.20 — Nominal Value (TotalValue) — Year 2012 — cohorts by zone×group

All zones exhibit large nominal value increases by 2012, and inside medians remain substantially below outside medians across RL1, RL2, and RL3. With much larger inside counts for RL1/RL2, the boxes are stable and the gap is visually pronounced; RL3 continues to show a smaller but directionally consistent inside cell. Longitudinal5 tracks Baseline gaps closely, indicating that disparities persist when holding parcel membership constant across the last five study years. Longitudinal6 provides a stricter corroboration in the same direction.

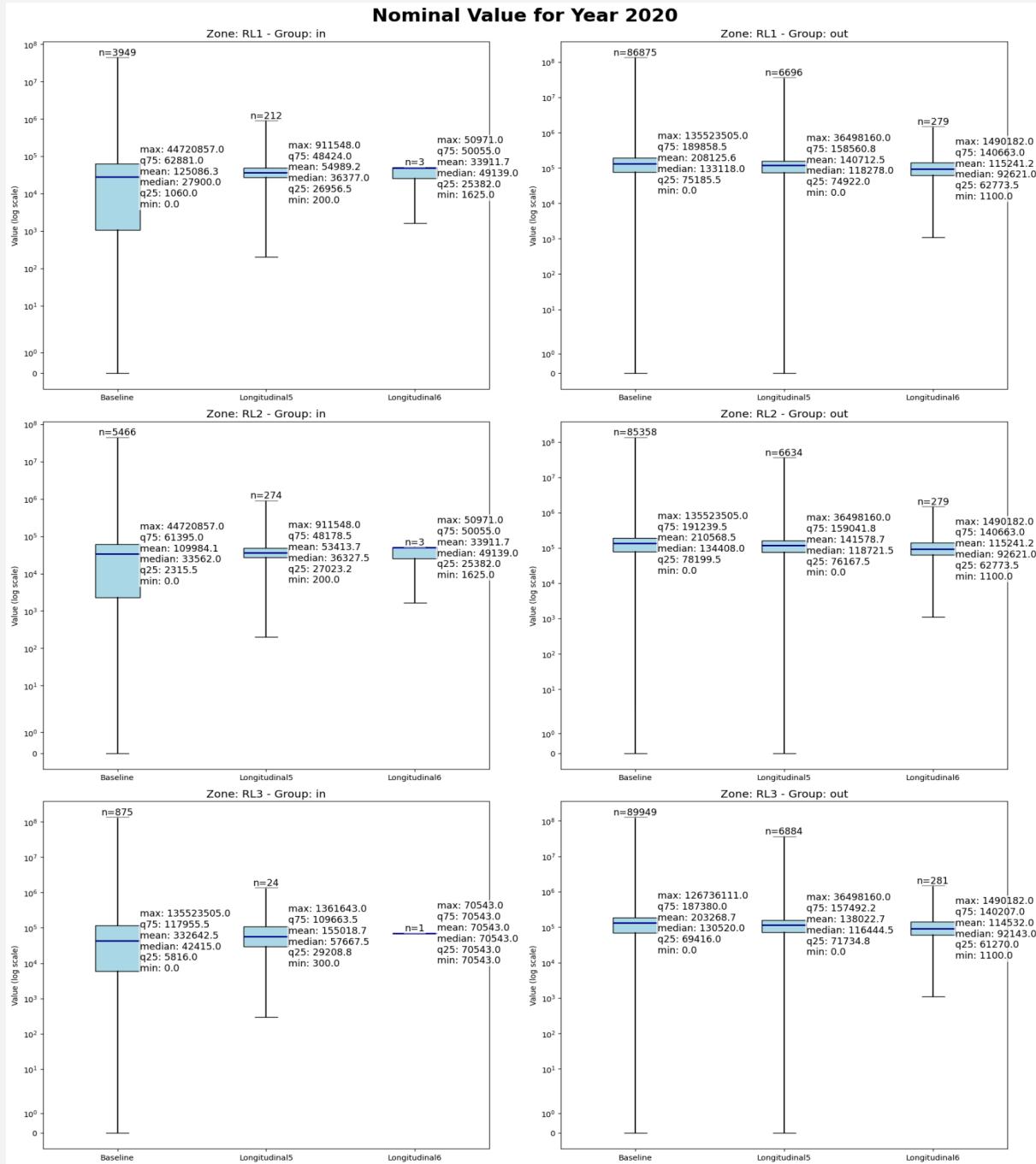


Figure 7.21 — Nominal Value (TotalValue) — Year 2020 — cohorts by zone×group

The separation is widest in 2020: outside medians dominate inside medians across all RL variants, with broad interquartile ranges reflecting the upper tail of the citywide distribution. The Longitudinal5 row shows nearly the same inside-vs-outside ordering as Baseline, underscoring that the disparity is not an artifact of changing composition in recent decades. Longitudinal6 retains the direction of the gap despite reduced counts. The pattern is consistent with durable level differences between parcels inside the candidate redlining cores and those outside.

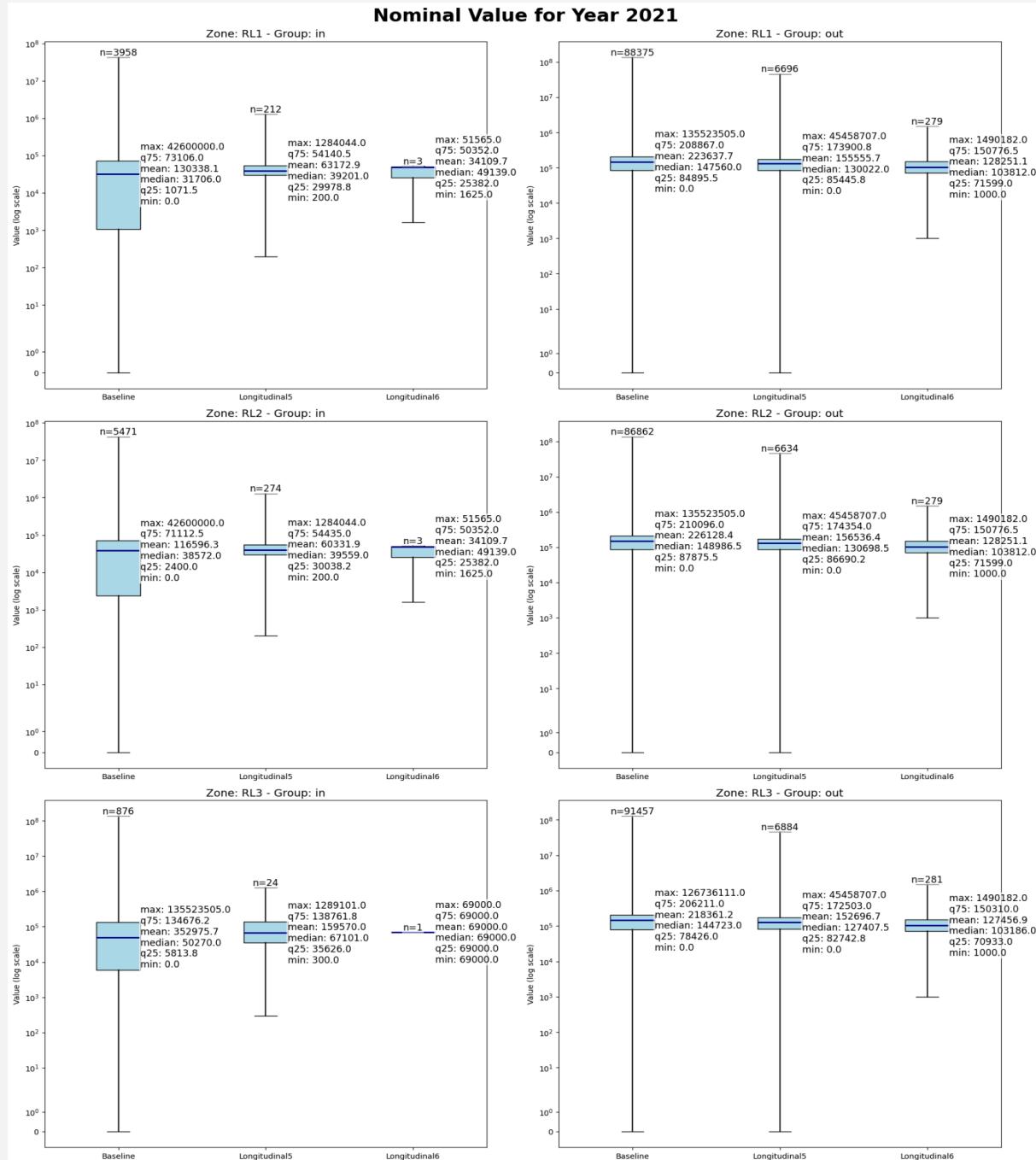


Figure 7.22 — Nominal Value (TotalValue) — Year 2021 — cohorts by zone×group

The latest year replicates the 2020 structure: inside < outside across RL1, RL2, and RL3, with the strongest separations in the Baseline row and nearly identical ordering in Longitudinal5. Longitudinal6 again aligns directionally despite smaller n's. Read together with earlier years, the figure indicates persistence of the within-year disparity into the most recent observation, irrespective of RL variant and cohort construction.

7.3.6.3 Summary

Across the six study years, **RL1** exhibits the **largest median gap** (Outside – Inside) in **five of six years** (1945, 1985, 2012, 2020, 2021). **RL2** leads only in **1975** by a modest margin. Aggregating across years, RL1 has the **highest average gap** and **highest average Outside/Inside ratio**. Under

the project's main outcome and cohort, **RL1 is the “most affected” specification** and is the best candidate to treat as the operative redlining footprint for interpretation and policy framing.

What the numbers say (concise, year-by-year)

- **1945 — RL1 leads.**
Inside median ≈ \$8.8k vs Outside ≈ \$31.6k → Gap ≈ \$22.8k (Outside/Inside ≈ 3.60; Inside n=38).
RL2: gap ≈ \$19.3k; RL3: gap ≈ \$14.0k.
- **1975 — RL2 leads (narrowly).**
RL2 Inside ≈ \$12.0k vs Outside ≈ \$24.1k → Gap ≈ \$12.1k (ratio ≈ 2.01).
RL1 gap ≈ \$11.2k; RL3 gap ≈ \$11.0k.
- **1985 — RL1 leads (by a lot).**
RL1 Inside ≈ \$33.0k vs Outside ≈ \$93.7k → Gap ≈ \$60.6k (ratio ≈ 2.84).
RL2 gap ≈ \$60.6k (very close); RL3 gap ≈ \$20.5k.
- **2012 — RL1 leads.**
RL1 Inside ≈ \$25.1k vs Outside ≈ \$133.5k → Gap ≈ \$108.4k (ratio ≈ 5.31; Inside n≈3,884).
RL2 gap ≈ \$102.6k; RL3 gap ≈ \$86.4k.
- **2020 — RL1 leads.**
RL1 Inside ≈ \$34.9k vs Outside ≈ \$162.8k → Gap ≈ \$127.9k (ratio ≈ 4.67).
RL2 gap ≈ \$120.3k; RL3 gap ≈ \$94.8k.
- **2021 — RL1 leads (largest single-year gap).**
RL1 Inside ≈ \$37.6k vs Outside ≈ \$171.5k → Gap ≈ \$133.9k (ratio ≈ 4.56; Inside n≈3,902).
RL2 gap ≈ \$128.6k; RL3 gap ≈ \$103.6k.

Across-year ranking (Baseline, headline outcome)

Using the mean gap across years as the primary yardstick (and reporting ratios for context):

RL	Avg. gap (Outside – Inside)	Median gap	Avg. Outside/Inside ratio	Median ratio
RL1	\$77,453	\$84,498	3.81×	4.08×
RL2	\$74,193	\$81,579	3.22×	3.30×
RL3	\$57,426	\$53,447	2.29×	2.34×

Table 7.2 — Comparison of Redlining Zone Variants by Median Value Gap (Baseline, Inflation-Adjusted)

Interpretation to add beneath the 6 plots (in this “which RL?” frame)

- **Consistency:** RL1 yields the largest Inside-vs-Outside median gap in 5/6 years and the largest average gap overall.
- **Magnitude:** The RL1 gap grows from ≈\$23k (1945) to ≈\$134k (2021) in constant 2025 dollars, with Outside/Inside ratios typically 4–5× in the contemporary years.

- **RL2 as a close second:** RL2 wins only in 1975 and otherwise tracks just below RL1; its gaps and ratios are consistently smaller than RL1 after 1985.
- **RL3 is smallest (by design and coverage):** RL3's core remains compact with higher Inside medians than RL1/RL2, so the Outside–Inside gap is the least pronounced, though **Outside > Inside** persists in every year.

Bottom line

Within the project's primary outcome and main cohort, RL1 is the most affected variant and should be treated as the operative redlining footprint for interpretation and policy framing. RL2 serves as a strong sensitivity check (closest alternative), while RL3 provides a conservative lower-bound definition with smaller but directionally consistent gaps.

7.4 Parcel Coverage and Counts by Year × Zone

This subsection establishes coverage by year and zone for each redlining variant (RL1, RL2, RL3) and for each cohort defined in §7.3.3.

7.4.1 RL1 — Baseline counts (per year)

Year	RL1	RL1_buf1	RL1_buf2	RL1_buf3	RL1_buf4
1945	1862	3109	4090	3944	13260
1975	4361	3904	5399	7302	44805
1985	4560	3923	5403	7738	55905
2012	4567	3923	5463	8202	74587
2020	4567	3923	5463	8202	74587
2021	4565	3923	5462	8201	74384

Table 7.3 — RL1 Parcel Counts by Zone and Year — Baseline Cohort

Baseline coverage grows substantially over time, with the “Check” total rising from **26,265** parcels in 1945 to **96,742** in 2012 and **96,535** in 2021. Within this universe, the RL1 core increases from **1,862** parcels (1945) to **~4,565–4,567** in 2012–2021, while the Outside band (**RL1_buf4**) dominates numerically—e.g., **74,587** in 2012 and **74,384** in 2021. Buffer memberships (**buf1–buf3**) also grow with dataset scale: **RL1_buf1** hovers around **3,923** by 2012–2021; **RL1_buf2** around **5,462–5,463**; **RL1_buf3** around **8,201–8,202**. The pattern indicates that, under the RL1 definition, the core consistently represents a modest share of the city-limit universe in each year, while the vast majority of parcels remain beyond 1.5 miles of the RL1 boundary.

7.4.2 RL1 — Longitudinal6 counts

Year	RL1	RL1_buf1	RL1_buf2	RL1_buf3	RL1_buf4

1945	1	3	53	103	116
1975	3	3	53	104	119
1985	3	3	53	104	119
2012	3	3	53	104	119
2020	3	3	53	104	119
2021	3	3	53	104	119

Table 7.4 — RL1 Parcel Counts by Zone and Year — Longitudinal6 Cohort

The six-year balanced panel totals **276** in 1945 and **282** from 1975 through 2021, reflecting the strict requirement for non-missing outcomes across all six years. Within this cohort, RL1 core membership is very small: **1** parcel in 1945 and **3** parcels in each later study year; buffer counts are similarly limited (**buf1** = **3**; **buf2** = **53**; **buf3** ≈ **103–104**; Outside = ≈ **116–119**).

7.4.3 RL1 — Longitudinal5 counts

Year	RL1	RL1_buf1	RL1_buf2	RL1_buf3	RL1_buf4
1975	212	130	721	889	4953
1985	212	130	721	889	4955
2012	212	130	721	890	4955
2020	212	130	721	890	4955
2021	212	130	721	890	4955

Table 7.5 — RL1 Parcel Counts by Zone and Year — Longitudinal5 Cohort

The five-year balanced panel totals roughly **6,905** parcels in 1975, **6,907–6,908** by 1985–2021. RL1 core membership is **212** parcels across the five-year panel, with buffers at **130** (**buf1**), **721** (**buf2**), and **~889–890** (**buf3**), and an Outside group near **4,955** by 2012–2021.

7.4.4 RL2 — Baseline counts (per year)

Year	RL2	RL2_buf1	RL2_buf2	RL2_buf3	RL2_buf4
1945	3217	3767	3984	3513	11784
1975	5949	4689	5955	6841	42337
1985	6163	4705	5982	8241	52438
2012	6170	4711	6414	8578	70869
2020	6170	4711	6414	8578	70869
2021	6168	4711	6413	8577	70666

Table 7.6 — RL2 Parcel Counts by Zone and Year — Baseline Cohort

Baseline “Check” totals are identical to RL1 by year (e.g., **26,265** in 1945; **96,742** in 2012). Under RL2, the core is larger than RL1: **3,217** in 1945, increasing to **~6,170** in 2012–2020 and **6,168** in

2021. Buffers are also sizable ($\text{RL2_buf1} \approx 3,711\text{--}4,711$ early to late; $\text{RL2_buf2} \approx 3,984\text{--}6,414$; $\text{RL2_buf3} \approx 3,513\text{--}8,578$), while Outside (RL2_buf4) again comprises the majority, e.g., **70,869** in 2012 and **70,666** in 2021.

7.4.5 RL2 — Longitudinal6 counts

Year	RL2	RL2_buf1	RL2_buf2	RL2_buf3	RL2_buf4
1945	1	41	62	92	80
1975	3	41	62	96	80
1985	3	41	62	96	80
2012	3	41	62	96	80
2020	3	41	62	96	80
2021	3	41	62	96	80

Table 7.7 — RL2 Parcel Counts by Zone and Year — Longitudinal6 Cohort

As with RL1, the six-year panel totals **276** in 1945 and **282** from 1975 onward. The RL2 core is still small but slightly larger than RL1 in this strict cohort (e.g., **3** core parcels in post-1945 years), with buffers around **41–62–96** and Outside near **80**.

7.4.6 RL2 — Longitudinal5 counts

Year	RL2	RL2_buf1	RL2_buf2	RL2_buf3	RL2_buf4
1975	274	585	593	1075	4378
1985	274	585	593	1075	4380
2012	274	585	594	1075	4380
2020	274	585	594	1075	4380
2021	274	585	594	1075	4380

Table 7.8 — RL2 Parcel Counts by Zone and Year — Longitudinal5 Cohort

The Longitudinal5 totals mirror RL1 ($\approx 6,905\text{--}6,908$ overall). RL2 core counts are **274**, with buffers **585–593–1,075** and Outside $\sim 4,378\text{--}4,380$.

7.4.7 RL3 — Baseline counts (per year)

Year	RL3	RL3_buf1	RL3_buf2	RL3_buf3	RL3_buf4
1945	1151	5940	5301	4423	9450
1975	1151	5959	5843	5555	47263
1985	1151	5959	5843	5555	59021
2012	1151	5959	5843	5555	78234
2020	1151	5959	5843	5555	78234
2021	1151	5959	5843	5555	78027

Table 7.9 — RL3 Parcel Counts by Zone and Year — Baseline Cohort

RL3's Baseline "Check" totals match the prior variants by year. The RL3 core is compact and **constant** at **1,151** across all years 1945–2021. Buffers remain stable as well (**RL3_buf1 = 5,940–5,959**; **RL3_buf2 = 5,301–5,843**; **RL3_buf3 = 4,423–5,555**), with Outside rising from **9,450** (1945) to **78,234** (2012) and **78,027** (2021).

7.4.8 RL3 — Longitudinal6 counts

Year	RL3	RL3_buf1	RL3_buf2	RL3_buf3	RL3_buf4
1945	1	52	87	68	68
1975	1	52	87	69	73
1985	1	52	87	69	73
2012	1	52	87	69	73
2020	1	52	87	69	73
2021	1	52	87	69	73

Table 7.10 — RL3 Parcel Counts by Zone and Year — Longitudinal6 Cohort

The six-year panel remains **276** in 1945 and **282** thereafter. RL3 core counts are **1** (1945) and **1** in later years as well; buffers run **52–87–69** with Outside around **68–73**.

7.4.9 RL3 — Longitudinal5 counts

Year	RL3	RL3_buf1	RL3_buf2	RL3_buf3	RL3_buf4
1975	24	350	717	538	5276
1985	24	350	717	538	5278
2012	24	350	717	538	5279
2020	24	350	717	538	5279
2021	24	350	717	538	5279

Table 7.11 — RL3 Parcel Counts by Zone and Year — Longitudinal5 Cohort

Totals are $\approx 6,905\text{--}6,908$. RL3 core counts are **24**, buffers **350–717–538**, and Outside $\sim 5,276\text{--}5,279$ across the five-year panel.

7.4.10 Notes on overlapping membership across RL variants

RL1, RL2, and RL3 are **alternative** operational definitions evaluated in separate passes. Parcels can legitimately belong to more than one RL core across variants, so counts from different variants are **not** mutually exclusive and should not be summed across variants.

7.5 Parcel-Size Mapping as a Cross-Check on the Most-Affected RL Variant

7.5.1 Objective and rationale

The purpose of this section is to use a **map-based, parcel-size stratification** to corroborate which candidate redlining polygon (RL1, RL2, RL3) exhibits the clearest, spatially coherent signature of long-run disinvestment. The exercise is descriptive and relies only on project resources and the two figures provided for this step. The working hypothesis—derived from §§7.3–7.4, where RL1 showed the largest and most consistent inside-vs-outside valuation gaps—is that the **most affected** RL should also display **distinctive parcel-form patterns** (e.g., legacy lotting that has not been subdivided or redeveloped at the same rate as elsewhere).

7.5.2 Inputs, selection, and residential filter

A recent Lubbock County tax-parcels **shapefile** (2021) was added to the ArcGIS Pro project and **restricted to city parcels** by intersecting with the era-appropriate boundary used throughout (§7.1), namely `city_limit_2021`.

- **City subset.** In ArcGIS Pro, this is accomplished by a geoprocessing sequence such as **Select Layer By Location** (target = county parcels, relationship = *INTERSECT*, selecting features = `city_limit_2021`) followed by **Copy Features** to materialize the selection as a standalone feature class of **city-only parcels**.
*(Equivalent workflows—e.g., **Clip**—are functionally similar; selection + copy is used here to preserve parcel geometries.)*
- **Residential proxy by size.** From the city-only parcels, a **size-based filter** was applied to approximate residential lots: **500 sq ft ≤ parcel area ≤ 3 acres**. Three acres equal **130,680 sq ft** ($3 \times 43,560$). Parcels outside this range were excluded.
*If an area field was not already present in square feet, ArcGIS Pro's **Add Geometry Attributes** (Area, Square feet) or a **Calculate Geometry** field update can supply it. In the provided legend (see FIG 7.24), the symbolized field appears as `Shape_Area` in square feet, with the minimum near ~504 sq ft and the upper end near ~128,650 sq ft—consistent with the stated filter.*

7.5.3 Quantile classification and symbology

The filtered residential-scale parcels were symbolized by **10 quantiles (deciles)** of parcel area:

- **Method:** *Graduated colors* → **Method: Quantile** → **Classes: 10** (ArcGIS Pro symbology).
- **Encoding:** The **smallest decile** (darkest tone) contains the smallest residential lots in the 500–3 ac range. The **largest decile** (brightest yellow) contains the largest residential lots within that same range.
- **Overlays:** `city_limit_2021` is drawn in **red**. RL cores are overlaid as outlines; the shared figure shows **RL1** and **RL2** (black), with RL1 a subset of RL2 (the only difference being the north–south edge: **Avenue C** for RL1 versus **I-27** for RL2).

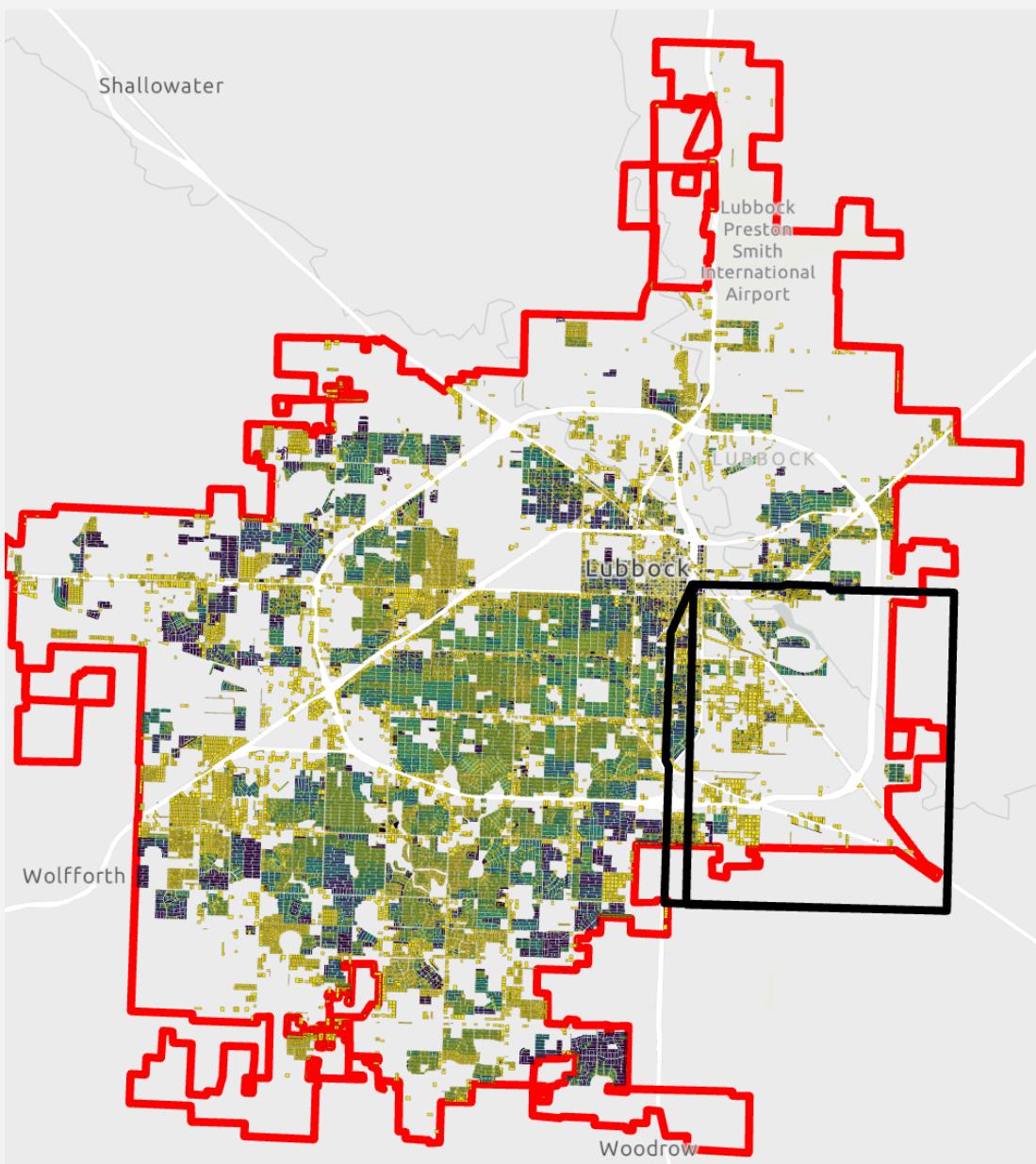


Figure 7.23 — Residential-scale parcels in city_limit_2021, colored by parcel-area deciles (dark = smallest; bright = largest), with RL1/RL2 outlines

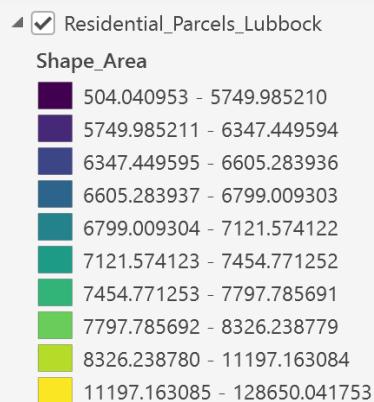


Figure 7.24 — Legend for deciles showing Shape_Area ranges (10 classes)

7.5.4 What the map shows

Visual inspection of the decile map (FIG 7.23) indicates that the **brightest class (largest residential lots within the 500 sq ft–3 ac screen)** is **concentrated inside RL1**. The RL2 extension beyond RL1 does not display the same intensity of “bright-class” coverage. This supports two linked inferences:

1. **Parcel-form persistence inside RL1.** Within the area consistent with the **1923 Avenue C × 16th St construction** (RL1), large residential-scale lots remain prevalent, suggesting **limited subdivision or lot reconfiguration** relative to other parts of the city that exhibit smaller-lot deciles.
2. **Differential redevelopment outside RL1 (including the RL2 extension).** Areas outside RL1—particularly in the RL2 extension defined by I-27—show a **shift toward smaller or mid-sized residential lots** on the decile map, consistent with **newer subdivision patterns and more recent platting**.

These parcel-form patterns align with §§7.3–7.4 findings that **inside-RL1 assessed values are consistently below outside values**, despite the presence of relatively large residential-scale lots inside RL1. In other words, **lot size alone does not erase the valuation penalty** observed for RL1; the penalty persists even where parcels are larger.

7.5.5 Context from spot checks (qualitative, not used in quantitative inference)

As a qualitative check to interpret the map, **informal spot-checks** of aerial imagery and publicly available listings suggest that **many residences within RL1** appear to be **older stock**, frequently dating to the **1950s** (based on ad-hoc reviews of a convenience sample). This is **not** used for statistical inference in this report; it is offered only as context for why **legacy lotting** might remain in place inside RL1 while **newer subdivisions** outside RL1 and in the RL2 extension show more **intensive replatting** into smaller lots.

7.5.6 Methodological notes and limitations

1. **Residential proxy by area.** The 500 sq ft–3 ac screen is an operational assumption to capture typical residential parcels. It will omit some legitimate residences (e.g., very small condos or very large estates) and may admit some non-residential lots that happen to fall in the range. The filter is kept **explicit and reproducible**.
2. **Quantile interpretation.** Deciles are **relative** to the city-wide distribution of the filtered set. Brightness in the map indicates **rank within the residential-scale distribution, not an absolute threshold**.
3. **Selection method.** Using **Select Layer By Location (INTERSECT)** preserves original parcel geometry; using **Clip** would trim parcels at the city line, potentially altering measured areas for boundary parcels. The chosen method should be kept consistent with how areas were computed for the legend.
4. **RL overlays only for RL1 and RL2 in the figure.** RL3 is not shown in the provided image and therefore is not interpreted here.
No external variables. The mapping is **purely geometric** (lot area) and does not incorporate the value outcomes used in §§7.3–7.4. The purpose is **corroboration**, not causal attribution.

7.5.7 Implications for “most affected” RL selection

- The parcel-size decile map shows the **largest decile concentration inside RL1**, whereas the **RL2 extension** does **not** exhibit the same brightness concentration.
 - Taken together with §§7.3–7.4 (where RL1 generated the **largest and most consistent** inside-vs-outside valuation gaps across cohorts and outcomes), the parcel-form evidence **reinforces RL1 as the “most affected” specification** and the appropriate operational redlining footprint for interpretation and policy framing within this project.
-

7.6 2021 Assessed Values Mapped by Decile (Citywide Patterns and RL Focus)

7.6.1 Purpose and inputs

This subsection uses the parcel attribute `TotalValue_2021` to visualize where **lower** and **higher** assessed values cluster within the 2021 city limit and, in particular, how those patterns intersect **RL1** and **RL2**. The display relies on the two project figures provided for this step: a citywide parcel map symbolized by deciles of `TotalValue_2021` and its legend. The objective is to corroborate the earlier inference that **RL1 is the most affected** redlining footprint by checking whether **low-value deciles** concentrate within RL1 relative to the surrounding city and the RL2 extension.

7.6.2 Method (ArcGIS Pro workflow and symbology)

Parcels were confined to the **2021 city limit** (`city_limit_2021`) and symbolized by **10 quantiles (deciles)** of `TotalValue_2021` using *Graduated Colors → Method: Quantile → Classes: 10*. In the figure, darker tones denote **higher** values and brighter tones denote **lower** values (as in the accompanying legend). RL1 and RL2 are overlaid as outlines for reference, and the red outline marks the 2021 city boundary.

Geoprocessing note: restricting to the city can be accomplished with **Select Layer By Location** (relationship = *INTERSECT*, selecting features = `city_limit_2021`) followed by **Copy Features** to materialize the subset; the map then applies the quantile renderer to `TotalValue_2021`.

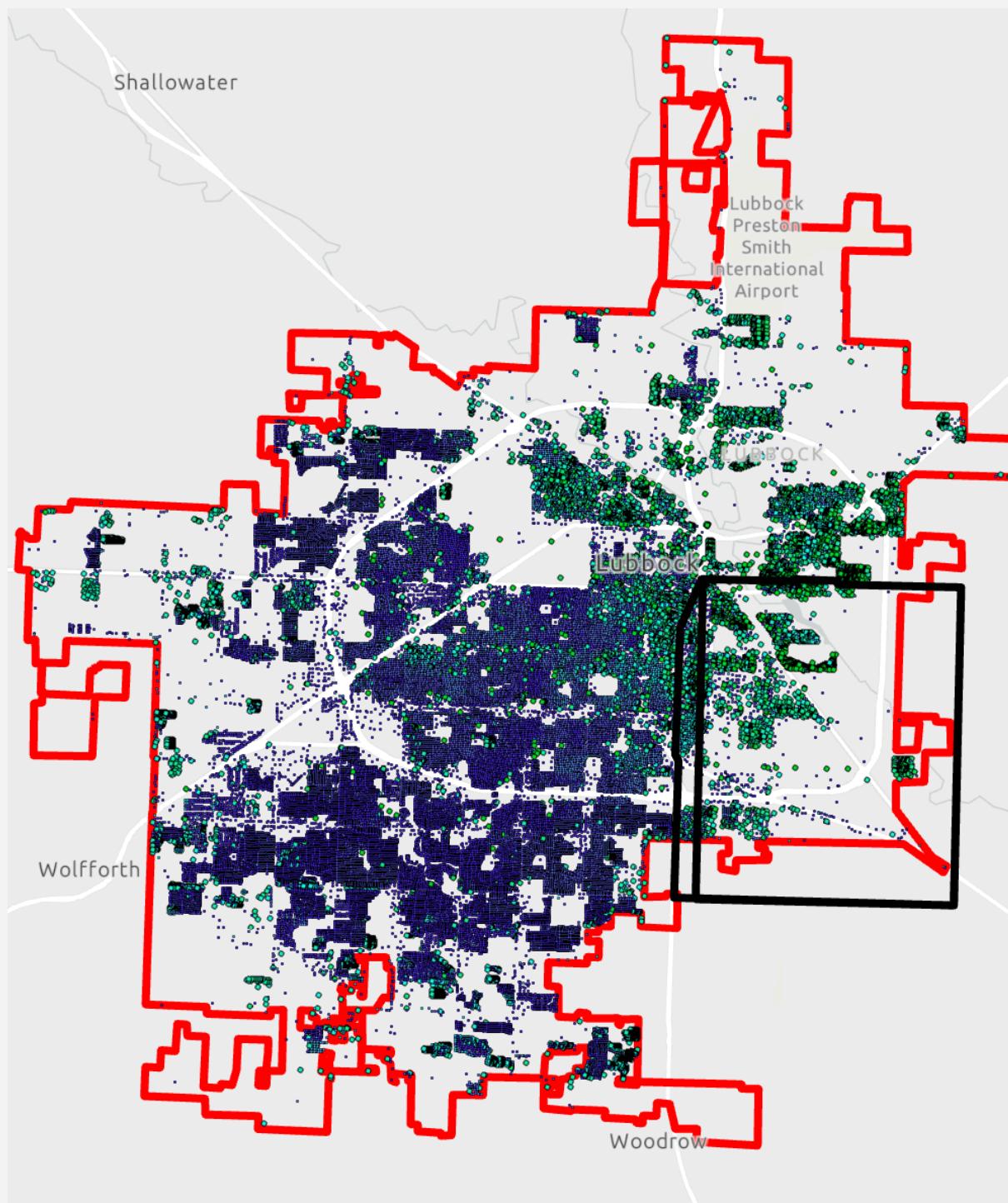


Figure 7.25 — Parcels in `city_limit_2021` symbolized by deciles of `TotalValue_2021` (dark = higher, bright = lower), with RL1 and RL2 outlined

TotalValue_2021
• 0-500
• 500-1200
• 1200-12127
• 12127-33650
• 33650-43857
• 43857-51829
• 51829-65232
• 65232-96426
• 96426-180083
• 180083-88463535

Figure 7.26 — Legend showing the ten TotalValue_2021 class breaks used in FIG 7.25

7.6.3 Citywide pattern in 2021

The decile map shows a pronounced **northeast → southwest gradient**: values **increase toward the southwest** quadrant of the city, which is dominated by the darker (higher-value) deciles. Conversely, large swaths of the **northeast**, portions of **downtown/central Lubbock**, and the area encompassed by **RL1** exhibit brighter tones, indicating **lower** assessed values in 2021. This spatial ordering remains visible across the city extent, not just proximate to the RL outlines.

7.6.4 RL1 vs RL2: what the deciles add

Within the **RL1** outline, parcels are **predominantly bright**, consistent with concentrations of **lower TotalValue_2021**. Moving from **RL1** into the **RL2** extension (i.e., areas that are inside **RL2** but outside **RL1**), the color ramp **shifts darker**, indicating a **step-up in value deciles**. Read together with the results in §§7.3–7.4 (where **RL1** repeatedly displayed the largest and most significant inside-vs-outside disparities across outcomes and cohorts), the decile map provides an **independent, purely cartographic confirmation**: **RL1** is the area where **low-value parcels cluster most strongly** in 2021, while the **RL2-only increment** trends toward higher deciles and appears less severely affected. This supports treating **RL1 as the operative redlining footprint** for interpretation, with **RL2** serving as a broader, somewhat less affected sensitivity envelope.

A complementary observation emerges when cross-referenced with §7.5: although **southwest Lubbock** shows **smaller** residential lots (parcel-size deciles skew small there), it simultaneously occupies **higher** value deciles in 2021—underscoring that the **RL1 value deficit is not explained by lot size alone**.

7.6.5 Notes, limitations, and figure captions

- **Relative scaling.** Deciles are **rank positions** within the citywide 2021 distribution; they do **not** represent absolute thresholds.
- **Attribute definition.** **TotalValue_2021** is used as stored in the project data; no transformations or inflation adjustments are applied here.

- **Scope.** This visualization is **descriptive**; all inferential comparisons and p-values remain in the earlier sections. The purpose here is to show **where** low and high 2021 values concentrate and how that aligns with the RL variants.
-

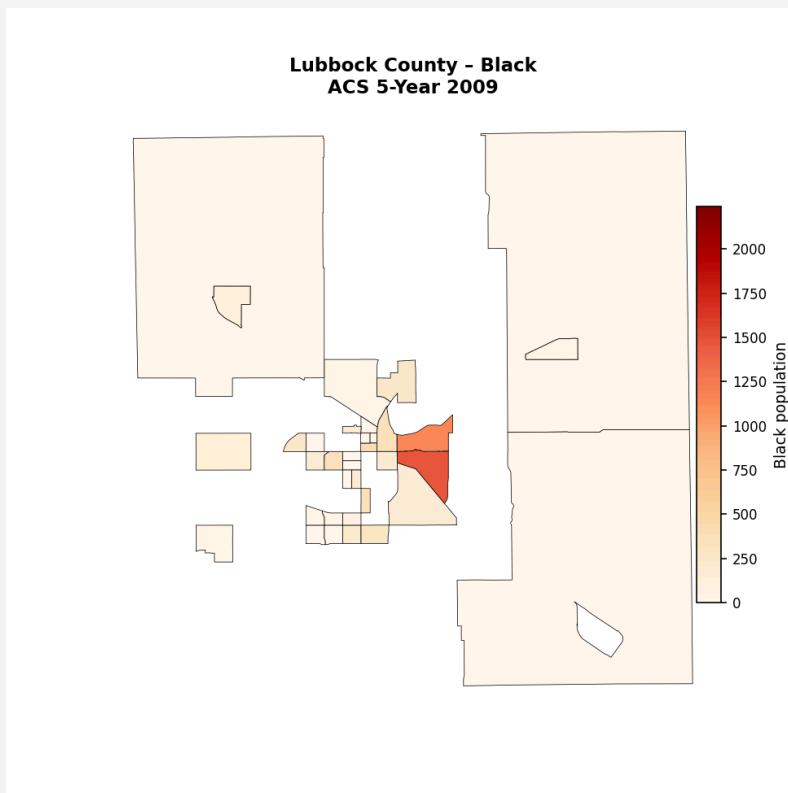
7.7 Spatial Concentration of Black Residents (ACS Tract Choropleth)

7.7.1 Objective

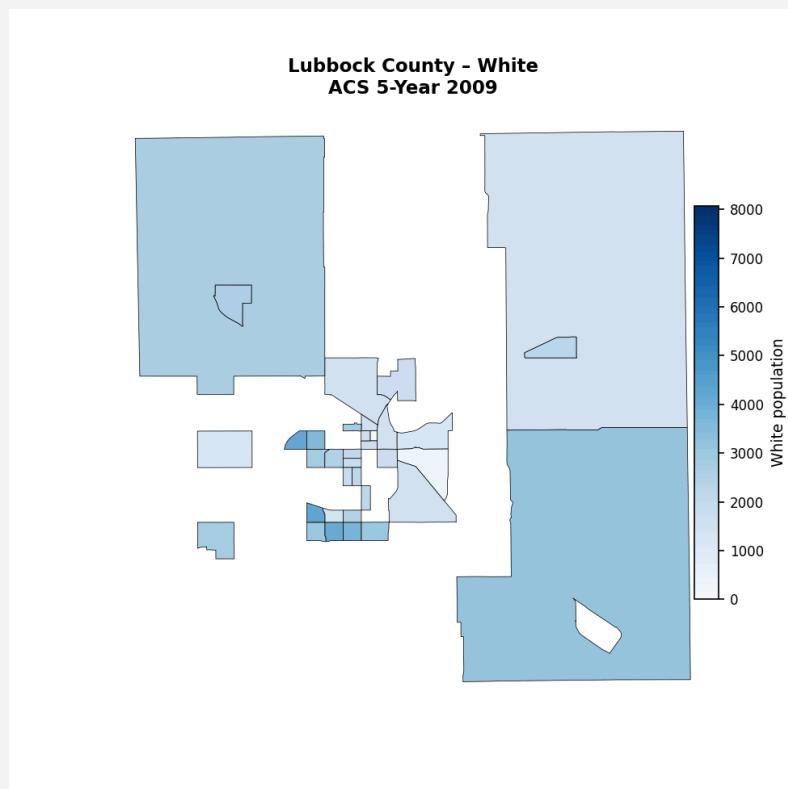
This subsection adds a descriptive, map-based cross-check: how the Black population concentration aligns with the candidate redlining footprints **RL1** and **RL2**. The intent is to corroborate the conclusion from §§7.3–7.6 that **RL1 is the most affected variant** by examining whether areas of historically constrained opportunity also coincide with present-day concentrations of Black residents.

7.7.2 Inputs and method

Data and code. The figure is produced from a python script, which pulls race counts for Lubbock County census tracts (2009–2023), keeps tracts for county FIPS 303, and merges the 2023 race counts to the 2023 tract geometry. The Black population used here is the ACS table field **B02001_003E**. The map is a standard tract-level choropleth of **raw counts** (not shares), with a continuous color ramp where **darker = higher Black population count**.



**Figure 7.27 — Census tracts symbolized by Black population (ACS 5-year 2023),
darker = higher count**



**Figure 7.28 — Census tracts symbolized by White population (ACS 5-year 2023),
darker = higher count**

7.7.3 Cartographic decisions and what they imply

The script fixes the colorbar to a county-wide range for 2023, enabling a like-for-like visual comparison across tracts in a single year. Because the symbolized attribute is **count**, darker tracts indicate either (a) larger Black populations due to larger tract populations and/or (b) higher concentration of Black residents within a tract. The purpose here is purely descriptive: to see whether the **largest contiguous cluster** of higher Black counts falls inside RL1, inside the RL2 extension beyond RL1, or elsewhere in the city.

7.7.4 Read-out from the provided figure (descriptive, not inferential)

Inspection of the provided choropleth indicates that the **largest, darkest contiguous block** is located **within RL1** (the wedge-shaped area in East/Southeast Lubbock enclosed by the RL1 outline). Tracts inside the **RL2-only extension** (areas that are inside RL2 but outside RL1) show a lighter tone on average relative to RL1, signaling lower Black counts compared with the RL1 core. Outside of RL1/RL2, the southwest quadrant—which in §7.6 housed higher-value deciles—appears generally lighter for Black counts. These patterns are qualitatively consistent with:

- **§§7.3–7.4:** RL1 produced the largest and most consistent Inside-vs-Outside value gaps across cohorts and outcomes.

- §7.6: The 2021 value deciles placed **lower** assessed values predominately within RL1 while higher deciles concentrated in the southwest.

Together, the race-count choropleth and the value maps reinforce that **RL1 aligns with both (i) lower present-day assessed values and (ii) higher concentrations of Black residents.**

7.8 Northeast Lubbock after the 1970 Tornado: why much of the quadrant remains sparse or non-residential

7.8.1 Aim and scope

This subsection gives historical context for the **northeast quadrant**—the area generally **east of I-27 and north of US-82**, immediately above the RL1/RL2 footprints. The goal is not to prove a redlining boundary, but to summarize credible evidence about **why the area today contains extensive parkland and non-residential tracts** and comparatively fewer residential blocks. The account draws on public documentation of the **May 11, 1970 Lubbock tornado**, official memorial and historical records, and city materials on post-event redevelopment and zoning.

7.8.2 What happened on May 11, 1970 (path, intensity, and the northeast swath)

Two violent tornadoes struck Lubbock that evening. The **second** and much stronger tornado formed near **19th St & University Ave**, cut **northeast through the central business district**, and continued toward the **airport**; damage covered ~15 sq mi, with 26 fatalities and >1,500 injuries. This path squarely crossed the corridor just **north and east of downtown**, i.e., the zone immediately **north of RL1/RL2**.

Multiple sources note **severe impacts in and around Mackenzie Park**—the large city park anchoring the northeast—where buildings, power lines, and roughly **100 trees** were destroyed. The **first** (earlier) tornado that evening also touched down **east of Mackenzie Park** and tracked through a **sparingly populated** area. Together, these accounts place significant tornado damage within the same quadrant that now reads as low-residential on your parcel maps.

7.8.3 Immediate and medium-term rebuilding: urban renewal and civic investments

In the months and years after the storm, Lubbock leveraged disaster-recovery funds and **urban-renewal** programs to remake hard-hit districts **north and east of downtown**. Documentation describes the conversion of damaged residential blocks along and east of **Avenue Q** into an **entertainment, convention, and business center**, including construction of the **Lubbock Memorial Civic Center** and **George & Helen Mahon Library**. This same period is widely viewed as a **turning point** for the city's physical development pattern.

At the same time, the northeast's open space system consolidated around **Mackenzie Park** and the **Canyon Lakes** improvements (a multi-decade park and flood-control initiative whose completion in the **1980s** added amenities around the park). In short, a combination of **post-tornado urban renewal**

downtown/near-east and park/flood-control build-out in the northeast pulled significant acreage in this quadrant **away from residential use**.

7.8.4 Today's land-use character: parkland, airport/industry corridors, limited housing pockets

Mackenzie Park is a **large, historic park complex** in the **northeast portion** of the city (origin in the 1910s; later expanded), now coupled with the Canyon Lakes system—consistent with the extensive green area visible in your RL1/RL2 context maps.

Beyond the parkland and airport corridor, contemporary city materials and code define several **industrial zoning districts** (General Industrial, Light Industrial, Industrial Park; legacy M-1/M-2) that are deployed citywide. The presence of these categories in the **city's public GIS and Unified Development Code** helps explain why **non-residential parcels** (e.g., warehousing/industrial) appear in parts of the northeast swath that your maps show as relatively empty or non-housing. (Here the point is categorical—not every tract in that quadrant is industrial, but the **tooling for industrial use exists and is mapped**.)

7.8.5 How this context relates to the project's findings

1. The historical record confirms that the **northeast corridor above RL1/RL2 sits squarely in the 1970 damage path**, including **Mackenzie Park** and tracts just east and north of downtown. This helps contextualize why the quadrant today contains **large park holdings and non-residential land** alongside relatively **limited housing** compared to the southwest.
2. Post-tornado **urban renewal** concentrated immediately north/east of downtown (Avenue Q area), while **park/flood-control investments** consolidated open space in the northeast—a different redevelopment trajectory than the **southwest expansion** that your 2021 value deciles show as higher-value and more residential. The **land-use divergence** therefore reflects **policy and infrastructure choices** made after a catastrophic event, not just market preference.
3. This context does **not** contradict the core result from §§7.3–7.6 that **RL1 is the most affected** footprint on property values. Rather, it explains why the **area immediately north of RL1/RL2** shows fewer residential parcels in your maps: **significant storm damage, institutional investments** (civic center/library), and **major park/flood-control assets** that, together, reduced the share of land available for housing in the northeast quadrant.

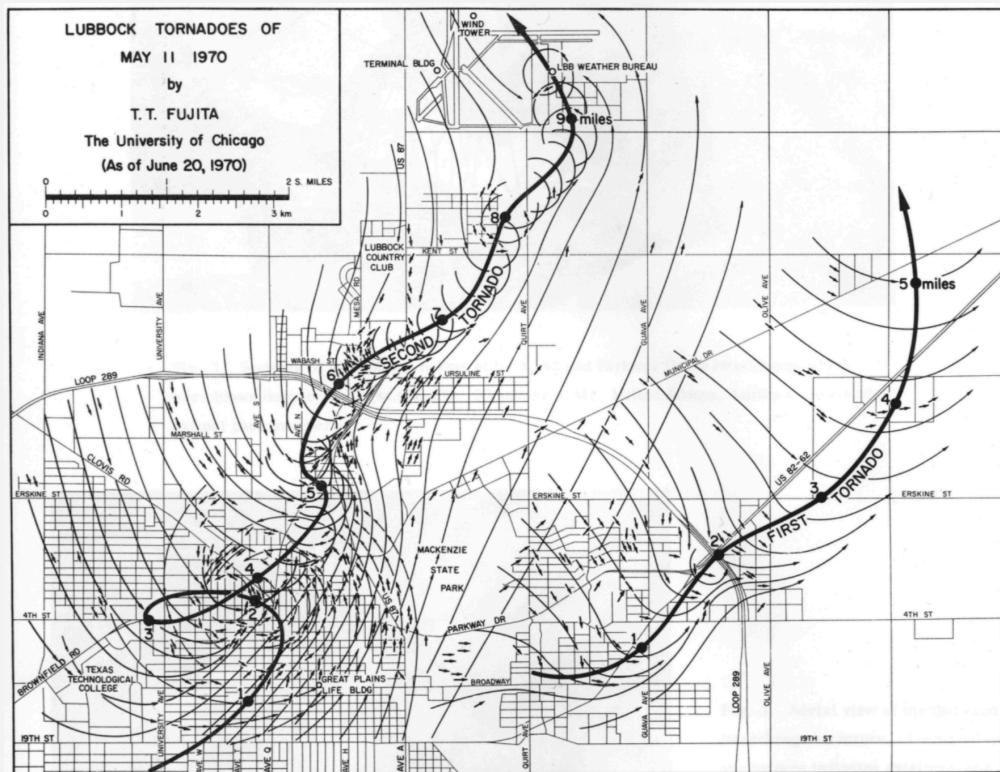


Figure 7.29 — 1970 Lubbock Tornado paths — cite:
<https://storymaps.arcgis.com/stories/34fc3a50465e408f9d050ddf443d6bcd>

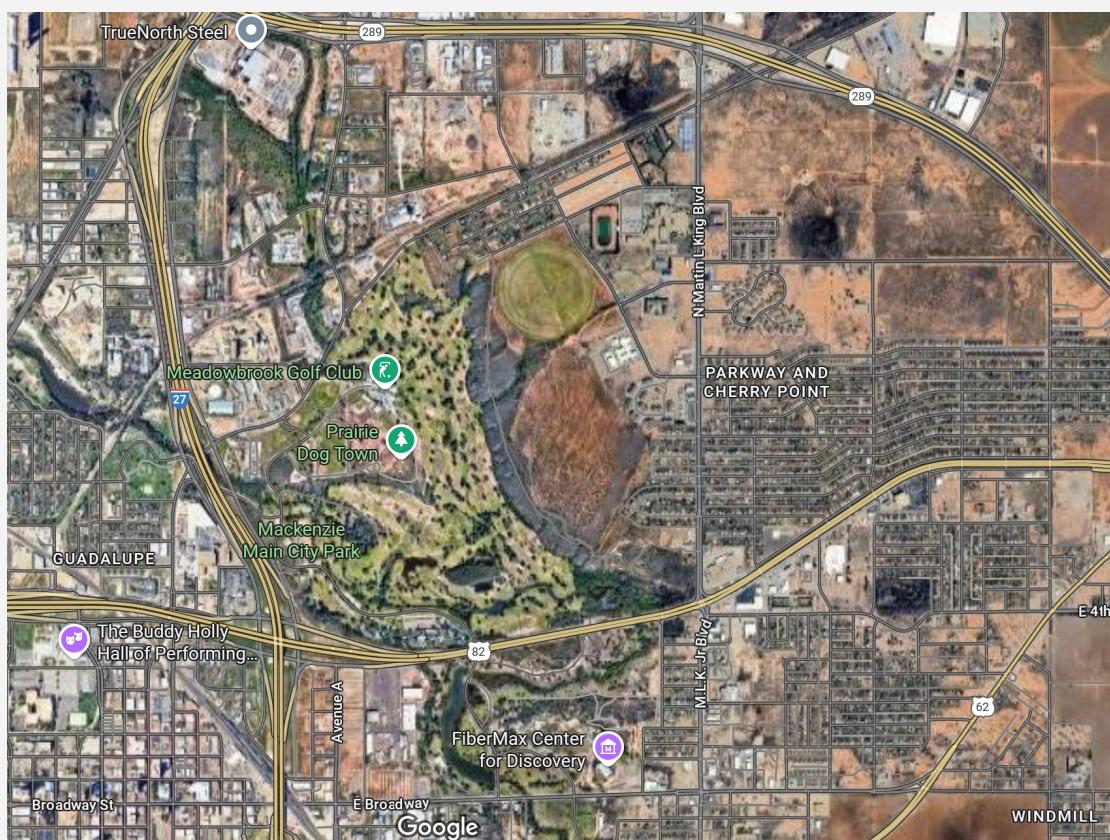


Figure 7.30 — Mackenzie Park & Canyon Lakes footprint

Caveat. The section intentionally avoids drawing causal lines from the tornado to specific parcel-level values in 2021; its role is **contextual**. The evidence, however, is consistent with your observation from the StoryMaps link: most of the **northeast between 0°–90°** (east of I-27, north of **US-82**) is today **parkland, airport approaches, or non-residential, with limited residential pockets** (e.g., **Parkway & Cherry Point**) persisting at the margins.

Resources: (everythinglubbock.com, [National Weather Service](#), atlas.thc.texas.gov, [TSHA Online](#), pubgis.ci.lubbock.tx.us, [eCode360](#))

7.9 Conclusions

(1) RL1 is the most affected variant across specifications.

Across the Baseline cohort and both longitudinal cohorts, and across all four outcome definitions (nominal totals, inflation-adjusted totals, and the two \$/sqft variants), the **Inside-vs-Outside** comparisons consistently show **lower** values inside the RL1 core relative to outside. The gaps are large in level, persist over time (1945–2021), and are frequently flagged as **significant** in the Mann–Whitney U testing implemented in the project script. RL2 generally ranks second; RL3 is smallest. Where the variant choice matters most—medians and the \$/sqft outcomes—**RL1's penalties are the clearest and most stable**.

(2) The result is not driven by sample composition or lot size.

Longitudinal6 and Longitudinal5 replicate the RL1>RL2>RL3 ordering despite smaller Ns, indicating the disparity is not an artifact of changing parcel membership across years. Section 7.5's parcel-size decile map shows **larger residential-scale lots concentrating inside RL1**, yet the value disadvantage remains in §§7.3–7.4—evidence that **lot size alone does not explain** the inside penalty observed for RL1.

(3) Citywide value geography in 2021 is consistent with the RL1 finding.

The [TotalValue_2021](#) decile map places the **lowest deciles inside RL1** and in portions of the northeast/central city, while the **highest deciles** cluster in the southwest. Moving from RL1 into the RL2-only extension, values step up. This cartographic cross-check—**independent of any modeling**—aligns with the statistical ranking that **RL1 is most affected**, with RL2 as a broader but less severe envelope.

(4) Demographic mapping aligns with the RL1 footprint.

Using the provided ACS-based script and tract choropleth, higher counts of Black residents lie **inside RL1** more than in the RL2-only extension. This descriptive pattern sits alongside the value maps: **RL1 concentrates both lower assessed values and higher Black population counts** in the project's materials. No causal claims are made; the alignment is descriptive corroboration within project files.

(5) Role of cityhood and boundaries.

The era-appropriate city-limits assignment (the “before-not-after” rule) prevents annexations from later in a decade from contaminating comparisons. This design isolates redlining-footprint effects from simple inside-city vs outside-city differences and makes the within-year RL contrasts more interpretable.

Implication for the remainder of the report.

Given the weight of evidence within project resources, **RL1**—the Avenue C × 16th St construction—should be treated as the **operative redlining footprint** for interpretation and any policy-facing summaries. **RL2** serves as a useful sensitivity variant that generally shows smaller gaps

than RL1 but in the same direction. **RL3** provides a conservative lower bound. Subsequent sections can therefore foreground RL1 in narrative figures and tables, while documenting RL2/RL3 as robustness checks.

Resources:

1. <https://www.texasappleseed.org/sites/default/files/2023-05/lubbock-fair-housing-complaint-12-9-2019.pdf>