Authors
Schlauch, Garrison
Greenhill, Simon

Publication Date
2026-09-10

Energy Institute WP 366

Data Centers Lower Nearby Housing Values

Garrison Schlauch
Simon Greenhill

September 2026

Data Centers Lower Nearby Housing Values

Garrison Schlauch*[a] and Simon Greenhill*[b,c]

[a]Department of Agricultural and Resource Economics, University of California, Berkeley, CA 94720.
[b]Center for the Environment, Harvard University, Cambridge, MA 02138.
[c]Department of Economics, Monash Business School, Monash University, Australia 3800.

September 2026

Abstract

Artificial intelligence is driving unprecedented investment in data centers, yet the local costs and benefits of these facilities remain largely unmeasured. We estimate how households value nearby data center development using a newly constructed database of U.S. data centers and housing transaction data. Our statistical analysis uses a matched stacked event study design that compares housing transactions near newly announced large data centers with transactions in observably similar areas near data centers that are announced later. Large data centers reduce housing values within 3 miles by 3% on average over the following 5 years. The effect is highly localized: prices within 1 mile fall by 7.2% on average, with smaller effects beyond 1 mile. Price declines persist for at least 5 years and are larger for hyperscale and high-capacity data centers, urban locations, and the first large data center in an area. Consistent with increasing public awareness of data center externalities, before 2021, prices responded primarily when data centers became operational, while for more recently announced data centers, the effect is capitalized almost entirely at announcement. Nationwide, these price declines imply $20.6 billion in housing wealth losses.

*Corresponding authors: Garrison Schlauch (gschlauch@berkeley.edu) and Simon Greenhill (sgreenhill@g.harvard.edu). The authors declare no competing interest. All errors are our own.

Annualized spending on U.S. data center construction exceeded $75 billion in July 2026, up 57% year-over-year and nearly sixfold from July 2022 (1). This investment has fueled rapid expansion: nearly 1,000 large data centers were announced between 2003 and 2025, with more than half since 2021 (Fig. S1), amid continued growth in cloud computing and, more recently, surging demand for artificial intelligence. Today, over 34 million Americans live within 3 miles of an operational or planned large data center.

Data centers generate externalities at every spatial scale. Recent work examines data centers' effects on greenhouse gas emissions, regional air quality, water quality and availability, and the electric grid (2–9). A parallel literature estimates economic effects at relatively aggregated geographic units such as counties, cities, and ZIP codes, focusing on labor market outcomes, business establishments, retail electricity prices, and housing prices (10–14). At more local scales, studies document environmental impacts including low-frequency noise (15, 16), increased temperatures (17, 18), and limited changes in air quality associated with co-located electricity generation (19). Residents near data centers have also raised concerns about nighttime lighting, construction, and visual disamenities (20).

Seventy percent of Americans oppose data center development near them, with 48–55% strongly opposed (21, 22). That opposition has grown rapidly: in August 2025, under a quarter of Americans strongly opposed local development; by May 2026, over half did (22). Residents near planned data centers have sought to block them through protests, legal complaints, and class action lawsuits (23–25), and some jurisdictions have responded by considering or enacting moratoria or outright bans (26).

Even as opposition mounts, many jurisdictions are seeking to attract data centers through tax incentives intended to generate local tax revenue and employment (27, 28). Evaluating these policies requires weighing those benefits against the costs borne by local residents and understanding how both are distributed within affected communities. Yet the net effect of data center development, as valued by nearby households, remains unmeasured.

This paper estimates how households value nearby data center development using the widely applied hedonic valuation method, which infers the value of local amenities and disamenities from how they are capitalized into housing prices (29, 30). We implement the hedonic method using what we believe to be the most comprehensive database of announcements and openings for large U.S. data centers from 2003 onwards (Fig. 1 panels A and C; 31). We focus on large data centers because their size makes them the most likely to generate measurable impacts. Following (32, 33), we define a large data center as a campus—a single facility or a collection of nearby related facilities—with at least 20 MW of power capacity or 100,000 square feet of building area.

We match the location and timing of large data center announcements with nearby residential housing transactions (Fig. 1 panels B and D). Because economic theory predicts that housing prices are forward-looking, we focus on announcements rather than openings in our main specification. Estimating dynamic effects allows us to trace how home prices respond following announcement and after a data center becomes operational. We also examine how the effect varies with distance and with characteristics of the data center and surrounding area.

To address omitted variable bias and estimate the causal effect of the announcement of a large data center on nearby housing prices, we use a matched stacked event study design (34). Because announcements occur at different times, conventional two-way fixed effects estimators may not recover an interpretable average treatment effect when effects vary across cohorts or over time (35, 36). We instead define each of the 346 announcements of large data centers in our main treatment sample as a separate sub-experiment with an event window spanning 5 years before through 5 years after the announcement (see Materials and Methods). For each sub-experiment, treated homes are those exposed to the announcement of their first large data center within 3 miles. We compare their transactions with those for not-yet-treated homes within 3 miles of a matched future large data center that remains unannounced through the end of the treated data center's event window. Fig. 1 panels A and C shows the locations and announcement timeline of the treated and comparison data centers in our main sample.

Using homes near future data centers as comparison units helps account for unobserved characteristics that may influence data center siting. To mitigate potential spillovers, candidate comparison data centers must also be at least 10 miles from any large data center that is announced or operational by the end of the event window. For each sub-experiment, we select the comparison data center candidate whose surrounding area most closely matches that of the treated data center in population density, median household income, share of non-White residents, and housing vacancy rate, with these characteristics measured over the same pre-treatment period defined relative to the treated data centers announcement. We then stack the 346 sub-experiments into a single estimation sample and estimate effects over the event window. This design ensures that transactions among treated homes are compared only with those among not-yet-treated homes and allows us to assess whether housing prices in the two groups follow similar trends before the treated data center's announcement.

Results

We find that the announcement of a large data center lowers housing values within 3 miles by 2.98% on average over the following 5 years (Fig. 2; 95% CI: [-4.59%, -1.38%]). We find no evidence of differential pre-trends: before the announcement, prices in the treatment and matched comparison groups move in parallel, consistent with the parallel trends assumption underlying our research design (see Materials and Methods). Values then decline by roughly 1.5% in the announcement year and a further 1.5% the following year. They continue to fall before stabilizing at about a 3.8% discount, indicating that the impact of data centers on home values persists at least through the medium term.

We attribute this event time profile to one measurement feature and one economic mechanism. First, because announcement timing is measured at an annual resolution, the estimate pools transactions in the months before and after the announcement during the announcement year, potentially attenuating the estimated effect for that year. Second, disamenities may become more salient to nearby homeowners once a data center is built and operating, increasing its capitalization into home values. Indeed, about 90% of operational data centers in our sample opened within 3 years of announcement (Fig. S2), around the point in time at which the estimated discount stabilizes. This pattern appears especially pronounced before the increased public awareness of data centers that accompanied the post-2020 hyperscale and AI boom, when the annual number of newly announced large data centers increased two- to threefold (Fig. S1). Before 2021, we find little price response to the announcement of a large data center, with most capitalization occurring once the data center becomes operational. In contrast, from 2021 onward, prices respond almost entirely at announcement, with virtually no additional decline after opening, consistent with increasing public awareness of data center externalities (Table S6).

Several disamenities potentially associated with data centers decay sharply with distance. For example, a case study in Prince William County, Virginia, examined data center noise—a leading source of local opposition (37)—and found that it was detectable up to 3 miles away (16). Visual disamenities and disruption from construction and increased traffic are likewise likely concentrated among the nearest homes.

Consistent with these channels, the estimated effect is concentrated within 1 mile of data centers on average (Fig. 3). Transactions within 1 mile show a 7.2% discount, nearly 2.5 times the pooled 3-mile estimate. Estimates beyond 1 mile are individually less precise and fall between -1.1% and -2.7%. We reject the joint null of no effects within 3 miles (p-value $<$ 0.01), including for the 2- to 3-mile range alone (p-value $= 0.04$), but fail to reject the joint null of no effects beyond 3 miles (p-value $= 0.27$).

We further investigate how the effect of data centers on housing values varies by data center and area characteristics (Fig. 4). Hyperscale data centers are associated with a substantially larger discount than other data center types, with imprecise effects for cryptomines. Data centers above the sample median power capacity of approximately 70 MW, which includes roughly 87% of treated hyperscalers, are likewise associated with larger discounts, consistent with disamenities that scale with size. Data centers in urban areas are associated with larger discounts than those in rural areas, where effects are smaller and imprecisely estimated. Finally, isolated data centers—those with no other large data center announced or operational within 10 miles before or during the event window—are associated with larger discounts, while additional data centers in the same area have a smaller marginal effect on home values.

Our results are robust across alternative research designs, specification choices, and sample restrictions. First, our results are robust to alternative approaches to specifying the research design, including using only repeated sales of the same homes, using three matched comparison groups instead of one, and weighting by the number of treated housing units in each sub-experiment (Fig. S3 and Table S4; 34). Second, the results are robust to alternative specification choices, including keeping outlier transactions, excluding controls for housing characteristics, and dropping transactions once homes are exposed to the announcement of another large data center (Table S5). The results also retain statistical significance when we two-way cluster the standard errors by census tract and calendar year (Table S5). Third, our findings are not driven by any particular state. In a leave-one-out analysis, we repeatedly re-estimate the main difference-in-differences specification (eqn. 2), each time excluding all sub-experiments whose treated data center is located in a given state. Each leave-one-out estimate is statistically different from zero at the 5% level and falls within the 95% confidence interval of the main estimate (Fig. S4). Lastly, we find no clear evidence that the announcement of a large data center affects housing sales (Fig. S6), suggesting that our estimated price effects are unlikely to be driven by large changes in transaction volume.

Our estimates imply an aggregate housing value loss of \$20.6 billion (see Materials and Methods for details). This is comparable to the \$24.5 billion loss attributed to wind turbine visibility nationwide (38), even though it is concentrated among far fewer homes.

Discussion

We find that the announcement of a large data center reduces housing values within 3 miles by about 3% on average. This effect is largest for homes within 1 mile of a data center and persists over a 5-year time horizon. Nationally, we estimate that large data centers have caused a \$20.6 billion reduction in housing values. Despite widespread discussion, speculation, and protest over these costs, they have not previously been quantified at scale. As a result, they generally have not been incorporated into data center siting decisions or analyses of the incentives jurisdictions use to attract data center development.

Although substantial for affected communities, these losses are modest relative to the scale of data center investment. Average construction costs reached \$10.7 million per MW in 2025 (39). Thus, a 100-MW data center—a scale that is expected to become the new norm (40)—can require more than \$1 billion in construction investment even before servers and other computing equipment are included, with some recent individual campuses seeing investments of \$25–27 billion (41, 42). These magnitudes suggest that compensating nearby homeowners for reductions in their home values may be feasible at modest expense relative to total project costs.

The larger effect we estimate for homes within 1 mile of a data center suggests that local disamenities such as noise and air pollution may be an important component. Yet sparse noise and air pollution monitoring near data centers makes it difficult to estimate the effects of on-site sources like cooling equipment and backup generators on noise and pollution levels (Fig. S7; see Supporting Information for details). While remotely sensed data (43) could provide an alternative means of studying the impacts of data centers on air pollution, they may be ill-suited for identifying the intermittent air pollution that backup generators are most likely to cause. Better understanding which local disamenities matter most is a crucial task for future research and policymaking, and could begin with increased environmental monitoring near data centers.

Data center development has been found to bring local benefits, including increases in employment and tax revenues (12, 13). These benefits should also be capitalized into housing prices to the extent that nearby homeowners value them, and thus our estimates net them out. More broadly, any benefits from the services supported by the computational power that data centers provide are likely shared by households in both the treatment and comparison groups and therefore are also netted out of our estimates. We thus find that within 7 miles of a large data center, capitalized local costs exceed capitalized local benefits over the 5-year period we study. This implies that our headline estimate understates the gross disamenity experienced by nearby households to the extent that they also benefit from data center development. At the same time, these locally borne costs are likely small relative to data centers' broader social benefits.

The negative effect of data center development on local housing values suggests opportunities for policy intervention. Policymakers may seek to compel developers to account for these local externalities through zoning and tax policy, or to mitigate them directly through measures such as noise barriers, visual screening, and setbacks from residential areas; our estimates provide a quantitative basis for doing so. A tax on a data center set to the expected capitalized loss across surrounding homes would price the local externality while allowing projects whose value exceeds that loss to proceed; blanket bans, by contrast, forgo the surplus from every project regardless of whether its benefits exceed local costs. Because our estimates capture only the benefits and costs borne by nearby homeowners, such a levy would be a lower bound on a data center's total external cost and would need to be combined with estimates of larger-scale damages such as air pollution and greenhouse gas emissions (9) to approach full internalization. These estimates could also inform compensation schemes for affected residents. Finally, because the discount attenuates sharply with distance, siting remains a first-order margin of adjustment: damages fall rapidly as data centers are placed farther from homes.

Materials and Methods

Data Center Data

We measure data center locations, characteristics, and development timing using a newly constructed database of U.S. data centers (31). The database combines facility-level information from three commercial datasets—Baxtel, Data Center Map, and TeleGeography—and Global Energy Monitor's open-source Cryptomining Tracker, all downloaded in 2025. Because each source includes data centers absent from the others, their combination substantially improves coverage. The resulting database contains 2,742 operational and planned campuses, which is 23–133% more than any of the three commercial datasets.

A data center campus comprises one or more related, geographically proximate facilities. Following (31), we group operators within the same facility, as well as facilities within 0.5 miles of one another operated by the same company, into a single campus. We aggregate facility characteristics within sources and then reconcile them across sources. When reconciling, we use the median across sources for power capacity and square footage and the maximum for building footprint area. Following recent work (31–33), we classify campuses as "large" if they have at least 20 MW of power capacity or 100,000 square feet of building area. For operational campuses missing square footage, the database uses Esri building footprint estimates derived from OpenStreetMap and Microsoft polygons; it then imputes missing power capacity from a log-log regression on campus square footage. These procedures identify 1,157 large operational or planned campuses. Because missing power capacity and square footage data are concentrated among smaller campuses, these imputations identify only two additional large campuses and provide greater confidence that we do not inadvertently exclude large campuses simply because the relevant size measures are missing.

The database also compiles announcement histories for large campuses. Because the commercial datasets do not report announcements, (31) manually searched company websites, press releases, news reports, public filings, and other contemporaneous sources to identify the first public announcement of the 994 large campuses that opened in 2003 or later or remained planned by the end of 2025. They identified announcement years directly for 751 of these campuses (75.6%) and, when unavailable, used construction start years for 116 campuses (11.7%). For 119 additional campuses (12.0%) that lack information on both milestones and were operational by the end of 2025, they imputed announcement years using the median announcement-to-opening interval among operational campuses of the same type: 1 year for cryptomining and 2 years for hyperscale and other types. The remaining eight campuses (0.8%) are planned but not yet operational and lack both announcement and construction start years. We exclude them because we cannot reliably infer their announcement years.

Housing Transaction Data

We use transaction-level residential housing data from version 3 of the CoreLogic Property and Owner Transfer files, covering sales from January 1998 through May 2024. These data provide parcel coordinates, sale dates, sale prices, and housing characteristics. We restrict the sample to residential housing transactions with non-missing values for sale price, sale date, total living area (square feet), number of bedrooms, and number of bathrooms. We then drop non-arm's-length transactions and those with sale prices below $1,000. Finally, we inflation-adjust sale prices to 2025 U.S. dollars, define the total number of rooms as the sum of bedrooms and bathrooms, and drop outlier transactions, which are defined as those below the 1st or above the 99th percentile of either sale price or total room count.

Demographic and Socioeconomic Data

We obtain pre-treatment demographic and socioeconomic characteristics for the treatment and comparison groups by spatially aggregating census block group (CBG)-level data from the 2000 Census and the 2005–2009 through 2017–2022 5-year American Community Survey (ACS) using publicly available data from IPUMS NHGIS (44). See SI Section A.1 for details on how we impute CBG-level data onto the treatment and comparison group areas.

Treatment Group Construction

Treatment is assigned at the home level based on the announcement of a large nearby data center. A home is treated by the first large data center announced within 3 miles. If a home is first exposed to an announcement by two or more large data centers in the same calendar year, we consider it treated by the closest one. Applying the restrictions described below yields a final treatment sample of 346 large data centers announced from 2003 through 2023, which we refer to as treatment events. For each treatment event, we retain transactions of treated homes from five years before through five years after the announcement.

To arrive at our final treatment sample, we begin with the 712 large data centers announced from 2003 through 2023. The sample begins in 2003, the earliest announcement year covered by (31), and ends in 2023 because the housing transaction data extend through May 2024, allowing us to observe transactions for at least 1 post-treatment year for each treated home. To avoid contamination from preexisting data centers, we exclude homes within 3 miles of any of the 159 large data centers announced or operational by 2002, as well as 10 data centers with unknown operational timing that may have been operating or announced by 2002. These exclusions leave 12 out of 712 data centers with no eligible homes within 3 miles, reducing the set of potential treatment events to 700.

We then assign homes to treatment sequentially by announcement year. Because homes exposed to an earlier announcement are no longer eligible for treatment by later announcements, some later-announced data centers have no previously untreated homes within 3 miles. This eliminates an additional 78 data centers, leaving 622 potential treatment events. Next, we drop 12 data centers in areas lacking the demographic and socioeconomic data required to select a matched comparison group. Finally, we exclude data centers without at least one transaction in the event study reference year (t=-1) and at least one transaction during the post-announcement event window (t=0 to 5). This eliminates many data centers in remote areas and those outside CoreLogics coverage. The final estimation sample contains 346 treatment events.

Comparison Group Construction

For each treatment event, we select a matched large data center announced more than 5 years later, or after May 2024, the final month of our housing transactions data. We impose the same transaction year requirements on potential matches as on the treatment group: at least one transaction within 3 miles in the reference year (t=-1) and at least one during the post-announcement window (t=0 to 5), with event time defined relative to the treated data center's announcement. To limit spillovers, candidate comparison data centers must also be at least 10 miles from any large data center announced or operational by the end of that window. The comparison group then consists of all transactions within 3 miles of the selected data center.

We select matches by minimizing the Mahalanobis distance between treated and potential comparison data centers using a parsimonious set of pre-treatment characteristics, measured relative to the treated data centers announcement: population density, median household income, share of non-White residents, and housing vacancy rate. For treatment events in 2010 or later, we measure these characteristics using ACS estimates covering the preceding 5 years. For events in 2009 or earlier, we use the 2000 Census because the earliest available 5-year ACS estimates extend into the post-treatment period. We match the 346 treatment events to 104 unique comparison data centers, as the same data center may serve as a comparison for multiple treatment events.

Statistical Analysis

Because data center announcements occur at different times, treatment adoption is staggered. Since conventional two-way fixed effects estimators may not recover an interpretable average treatment effect in such settings (35, 36), we use a stacked event study, which ensures that transactions among treated homes are compared only with those for homes that remain untreated through the relevant event window (34). Although several alternative estimators accommodate staggered adoption and heterogeneous treatment effects (35, 45–47), the stacked framework is particularly well suited to our setting because it makes it straightforward to define each of our 346 treatment events as a separate sub-experiment and pair it with the single clean comparison group selected through our matching procedure above.

To construct the stacked dataset, we create a separate panel for each sub-experiment containing the treated homes, their matched comparison homes, and their transactions occurring from 5 years before through 5 years after the treated data center's announcement. We define event time relative to that announcement and concatenate the 346 sub-experiment-specific panels into a single estimation sample. Using this stacked dataset, we estimate the following event study specification via ordinary least squares:

$$\log(price_{ige\tau t}) = \sum_{\substack{h=-5,\ldots,5 \\ h\neq -1}} \beta_h \left(\mathbf{1}\{\tau = h\} \times Treat_{ige}\right) + \alpha_{ge} + \delta_{e\tau} + \boldsymbol{X}'_{ige\tau t}\gamma + \epsilon_{ige\tau t}. \tag{1}$$

The outcome, $\log(price_{ige\tau t})$, is the natural logarithm of the sale price of home i assigned to treatment or comparison group g in sub-experiment e sold in event year τ on calendar date t. The variables $\mathbf{1}\{\tau = h\}$ are indicators for the event year relative to treatment event e. These indicators are interacted with $Treat_{ige}$, an indicator equal to one for homes in the treatment group of sub-experiment e and zero otherwise. Following standard practice, we omit event year $\tau = -1$ from this set of interactions.

The term α_{ge} denotes group-by-sub-experiment fixed effects, allowing the treatment and comparison groups in each sub-experiment to have separate intercepts. These fixed effects absorb time-invariant differences across groups within a sub-experiment, including baseline differences in housing prices. The term $\delta_{e\tau}$ denotes sub-experiment-by-event year fixed effects. These fixed effects absorb shocks that are common to treatment and comparison group homes within the same sub-experiment and event year. For example, they control for changes in national housing market conditions and other macroeconomic factors that affect both groups. Together, these two sets of fixed effects imply that identification comes from comparing how housing prices evolve in the treatment group relative to the matched comparison group within each sub-experiment.

The vector $\boldsymbol{X}_{ige\tau t}$ contains observable characteristics for each home i: square footage and the number of bedrooms plus bathrooms. Finally, $\epsilon_{ige\tau t}$ is an unobserved error term. We cluster standard errors at the Census tract level to allow for arbitrary serial correlation in housing prices within local housing markets both within and across sub-experiments. Because Census tract boundaries change over time, we use 2000 Census tracts throughout the analysis.

The coefficients of interest are the event study parameters β_h. Each β_h measures the difference in log sale prices between the treatment and comparison groups h years relative to the announcement, normalized to the corresponding difference in the omitted pre-announcement year ($h = -1$). Because treatment is defined by the initial announcement of a nearby data center, β_h captures the capitalization of information revealed at announcement together with subsequent effects from the announced data centers construction, operation, and potential expansion, as well as from any additional nearby data center development during the event window.

The parameters β_h identify the causal effect of the announcement of a large data center on local housing prices under two main assumptions: no anticipation and parallel trends. The no anticipation assumption requires that housing prices do not respond before the assigned treatment year, which is plausible given that buyers and sellers are unlikely to learn of a data center before it is publicly announced or under development. The parallel trends assumption requires that, absent data center announcement, housing prices in treatment and comparison groups would have evolved similarly.

While these assumptions cannot be tested directly, our research design reduces the likelihood that they are violated. Comparison groups are drawn from homes surrounding future data center sites, making them more likely to share unobserved determinants of data center siting with the corresponding treatment groups, and are matched on pre-treatment demographic and socioeconomic characteristics. This design choice makes our treatment and control groups potentially more likely to follow parallel trends in their housing prices (34). In addition, we report pre-treatment event study coefficients. Finding that these coefficients are economically small and statistically indistinguishable from zero is consistent with both parallel trends and no anticipation effects.

To summarize the average post-announcement effect, we also estimate the following stacked difference-in-differences regression:

$$\log(price_{ige\tau t}) = \beta\left(Treat_{ige} \times Post_{\tau}\right) + \alpha_{ge} + \delta_{e\tau} + \boldsymbol{X}'_{ige\tau t}\gamma + \epsilon_{ige\tau t}, \tag{2}$$

where $Post_{\tau} = \mathbf{1}\{\tau \geq 0\}$ is an indicator equal to one in the announcement year and thereafter, and zero in the pre-announcement years. The coefficient β measures the average effect of the announcement of a large data center on treated homes over the 5-year post-announcement event window.

To examine how treatment effects vary with proximity to data centers, we estimate

$$\begin{aligned}\log(price)_{ige\tau t} = &\sum_{b\in\mathcal{B}} \beta_b \left(Treat_{ige} \times Bin^b_{ige} \times Post_{\tau}\right) \\ &+ \sum_{b\in\mathcal{B}\setminus b_0} \lambda_b \left(Treat_{ige} \times Bin^b_{ige}\right) + \alpha_{ge} + \delta_{e\tau} + \boldsymbol{X}'_{ige\tau t}\gamma + \epsilon_{ige\tau t},\end{aligned} \tag{3}$$

where Bin^b_{ige} is an indicator equal to one if treated home i in group g of sub-experiment e lies within distance bin b of the data center, and zero otherwise. Because the group-by-sub-experiment fixed effects, α_{ge}, already include a common baseline intercept for all treated homes within each sub-experiment, we omit one lower-order interaction to avoid multicollinearity. The remaining lower-order interactions allow treated homes in different distance bins to have different pre-announcement price levels. The coefficients β_b therefore measure the difference-in-differences treatment effect for homes in distance bin b relative to the full matched comparison group. Estimating the distance bin effects in a single regression also allows us to conduct joint hypothesis tests across distance bins.

We use the same comparison group for every treated distance bin rather than introducing additional interactions that partition the comparison group. This common reference group allows differences across the estimated β_b coefficients to be interpreted as differences in treatment effects by distance rather than differences in the comparison group. We extend the treated distance bins through 7 miles. Candidate comparison data centers are at least 10 miles from the treated data center, and the comparison group contains all transactions among homes within 3 miles of the selected comparison; thus, 7 miles is the maximum treated radius that cannot overlap the matched comparison group. To avoid conflating the distance gradient from the original treatment event with exposure to the announcement or development of a subsequent nearby data center, we exclude transactions for treated homes once another large data center is announced within 7 miles, while retaining all prior transactions.

Lastly, to examine heterogeneous effects by data center and area characteristics, we estimate

$$\log(price_{ige\tau t}) = \sum_{c} \beta_c \left(Treat_{ige} \times Post_{\tau} \times C_e^c\right) + \alpha_{ge} + \delta_{e\tau} + \boldsymbol{X}'_{ige\tau t}\gamma + \epsilon_{ige\tau t}, \tag{4}$$

where C_e^c is a mutually exclusive and exhaustive set of indicators classifying each sub-experiment e according to a characteristic of interest. The coefficients β_c therefore measure the average treatment effect for treated homes for each characteristic.

Damages Estimation

We estimate aggregate housing wealth losses by multiplying the estimated proportional effect of the announcement of a large data center on housing prices by the average pre-treatment sale price of treated homes in our sample and the estimated total number of pre-treatment treated homes across all sub-experiments. Formally, we calculate damages as $D = -\widetilde{\beta} \times \overline{P} \times N$, where $\widetilde{\beta} = \exp(\widehat{\beta}) - 1 = -0.0298$ is the proportional effect on housing prices obtained by rescaling the log-point difference-in-differences estimate $\widehat{\beta}$ from eqn. (2), $\overline{P} = 387{,}380$ is the average pre-treatment sale price of treated homes in 2025 dollars, and $N = 1{,}780{,}509$ is the estimated number of pre-treatment homes in treated areas, measured using Census and ACS data (See Section A.1 for more details). Using the Census-based housing stock, rather than only homes observed to transact in CoreLogic, accounts for incomplete transaction coverage and homes that do not transact during the 5-year pre-period. This calculation yields an estimated aggregate housing wealth loss of approximately $D = -(-0.0298) \times 387{,}380 \times 1{,}780{,}509 \approx \$20.6$ billion.

Acknowledgments

We are particularly grateful to Michael Anderson and Maximilian Auffhammer for their feedback and support. We also thank Severin Borenstein, Tamma Carleton, Lucas Davis, Meredith Fowlie, James Sallee, Joseph Shapiro, Reed Walker, Paige Weber, and seminar participants at UC Berkeley for helpful comments. Claudia Escorcia, Diego Flores, Mariah Gutierrez, Satvika Iyer, Tony Lomeli, Kiana Lovgren, Duong Nguyen, Danicka Sailer, Sarah Siemsgluess, and Aditi Swaminathan provided excellent research assistance. The data centers database used in this paper was built in collaboration with Coly Elhai, Kevin Butler, Gianluca Guidi, and Falco J. Bargagli-Stoffi.

References

[1] U.S. Census Bureau, Value of construction put in place survey: Private construction spending time series, seasonally adjusted (July 2026 release, published 1 September 2026) (2026) Accessed 2 September 2026.

[2] A Papp, D Almond, S Zhang, Bitcoin and carbon dioxide emissions: Evidence from daily production decisions. *Journal of Public Economics* **227**, 105003 (2023).

[3] G Guidi, et al., Environmental burden of United States data centers in the artificial intelligence era. *arXiv preprint arXiv:2411.09786* (2024).

[4] C Elsworth, et al., Measuring the environmental impact of delivering AI at Google scale (2025) arXiv:2508.15734 [cs.AI].

[5] G Guidi, F Dominici, The hidden water geography of US hyperscale data centers in the AI era. *arXiv preprint arXiv:2607.02531* (2026).

[6] D Hernandez-Cortes, K Meng, P Weber, The environmental costs and geography of US data center expansion. (2026).

[7] O Kay, R Reaser, R Taylor, Processing power: The effect of data centers on wholesale electricity markets. *Federal Reserve Bank of Dallas, Working Papers* **2026** (2026).

[8] J Mamkhezri, X Sun, Y Yang, The hidden cost of the cloud: Data centers and electricity market inefficiency. (2026).

[9] N Muller, Measuring the impact of data centers in the United States economy: Monetary damage from air pollution and greenhouse gas emissions, (National Bureau of Economic Research), Technical report (2026).

[10] M Benetton, G Compiani, A Morse, When technology processors come to town: High electricity use spillovers to the local economy, Technical report (2025).

[11] A Priest, Not in my back yard! the effects of data centers on housing prices. *The Effects of Data Centers on Housing Prices (February 27, 2026)* (2026).

[12] FE Alvarez, D Argente, J Chow, D Van Patten, Data centers and local economies in the age of AI: A shiftshare approach (2026).

[13] Y Zeng, D Yue, The Local Economic Effects of AI Data Center Entry. *Academy of Management Proceedings* **2026**, 15909 (2026).

[14] R Meeks, J Pless, Z Qi, Z Wang, Who pays for growth? Evidence from data centers and the grid. (2026).

[15] JW Lyver, Data center noise study for Prince William, Fauquier, and King George Counties and the Town of Warrenton, Technical report (2022).

[16] JW Lyver, Noise analysis for PW County data centers briefing to DCOAG & JMT, Technical report (2024).

[17] DJ Sailor, SS Abolhassani, EP Martin, Data center waste heat as an emerging urban thermal hazard: First field measurements of neighborhood-scale air temperature impacts. *ASME Journal of Engineering for Sustainable Buildings and Cities* **7**, 024501 (2026).

[18] A Marinoni, et al., The data heat island effect: Quantifying the impact of AI data centers in a warming world. *arXiv preprint arXiv:2603.20897* (2026).

[19] C Jia, Air quality analysis reveals minimal changes after xAI data center opens in pollution-burdened Memphis neighborhood. *The Conversation* (2025) Check for a peer-reviewed version before citing.

[20] C Smith-Schoenwalder, 'Living in hell': Data center neighbors grapple with noise, air pollution. *U.S. News & World Report* (2026).

[21] J Jeffrey, Americans oppose AI data centers in their area. *Gallup.com* (2026) Section: Politics.

[22] R Meyer, Americans now overwhelmingly oppose new data centers near them. *Heatmap* (2026).

[23] A Wittenberg, Anger over data centers is fueling a new political movement. *POLITICO* (2026).

[24] Citizens Concerned About Wolf Hollow, Citizens Concerned About Wolf Hollow v. Marathon Digital Holdings, Inc. (Verified Petition, Application for Permanent Injunction, and Request for

Disclosure, Cause No. C2024253, 335th Judicial District Court, Hood County, Texas) (October 4, 2024).

[25] Haley et al., Haley et al. v. xAI Corp., Space Exploration Technologies Corp., and MZX Tech LLC (Complaint, No. 3:26-cv-00148-MPM-RP, N.D. Miss.) (June 8, 2026).

[26] National Conference of State Legislatures, Which states are banning data centers? (2026).

[27] Stream Data Centers, Tax incentives for data centers (2026).

[28] A Gargano, M Giacoletti, Subsidizing the cloud: US state incentives to data centers. *Available at SSRN 5881105* (2025).

[29] S Rosen, Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of political economy* **82**, 34–55 (1974).

[30] HS Banzhaf, Difference-in-Differences Hedonics. *Journal of Political Economy* **129**, 2385–2414 (2021).

[31] G Schlauch, et al., Building and analyzing a comprehensive database of U.S. data centers (2026) Working paper. Draft available upon request.

[32] Uptime Institute, Best-in-class data center provisioning, (Uptime Institute), Keynote report 36 (2023).

[33] S Ghosh, J Damodara, Scalable data centers–power generation and delivery challenges and solutions. *Energy Reports* **15**, 109178 (2026).

[34] C Wing, SM Freedman, A Hollingsworth, Stacked difference-in-differences, (National Bureau of Economic Research), Technical report (2024).

[35] C De Chaisemartin, X dHaultfoeuille, Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review* **110**, 2964–2996 (2020).

[36] A Goodman-Bacon, Difference-in-differences with variation in treatment timing. *Journal of econometrics* **225**, 254–277 (2021).

[37] J Holzman, I spent the day at a noisy data center. Here's what I learned. *Heatmap News* (2026).

[38] W Guo, L Wenz, M Auffhammer, The visual effect of wind turbines on property values is small and diminishing in space and time. *Proceedings of the National Academy of Sciences* **121**, e2309372121 (2024).

[39] JLL, 2026 global data center market outlook (2026) Accessed August 23, 2026.

[40] CBRE, Data centers: U.S. real estate market outlook 2025 (2025) Accessed August 23, 2026.

[41] Meta, Meta announces joint venture with funds managed by Blue Owl Capital to develop Hyperion data center (2025) Accessed September 6, 2026.

[42] Vantage Data Centers, Vantage Data Centers unveils plans for "frontier," a $25b mega-campus in Texas to meet unprecedented AI demand (2025) Accessed September 6, 2026.

[43] A van Donkelaar, et al., Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology* **55**, 15287–15300 (2021).

[44] J Schroeder, et al., IPUMS national historical geographic information system: Version 20.0 [dataset] (2025).

[45] B Callaway, PH SantAnna, Difference-in-differences with multiple time periods. *Journal of econometrics* **225**, 200–230 (2021).

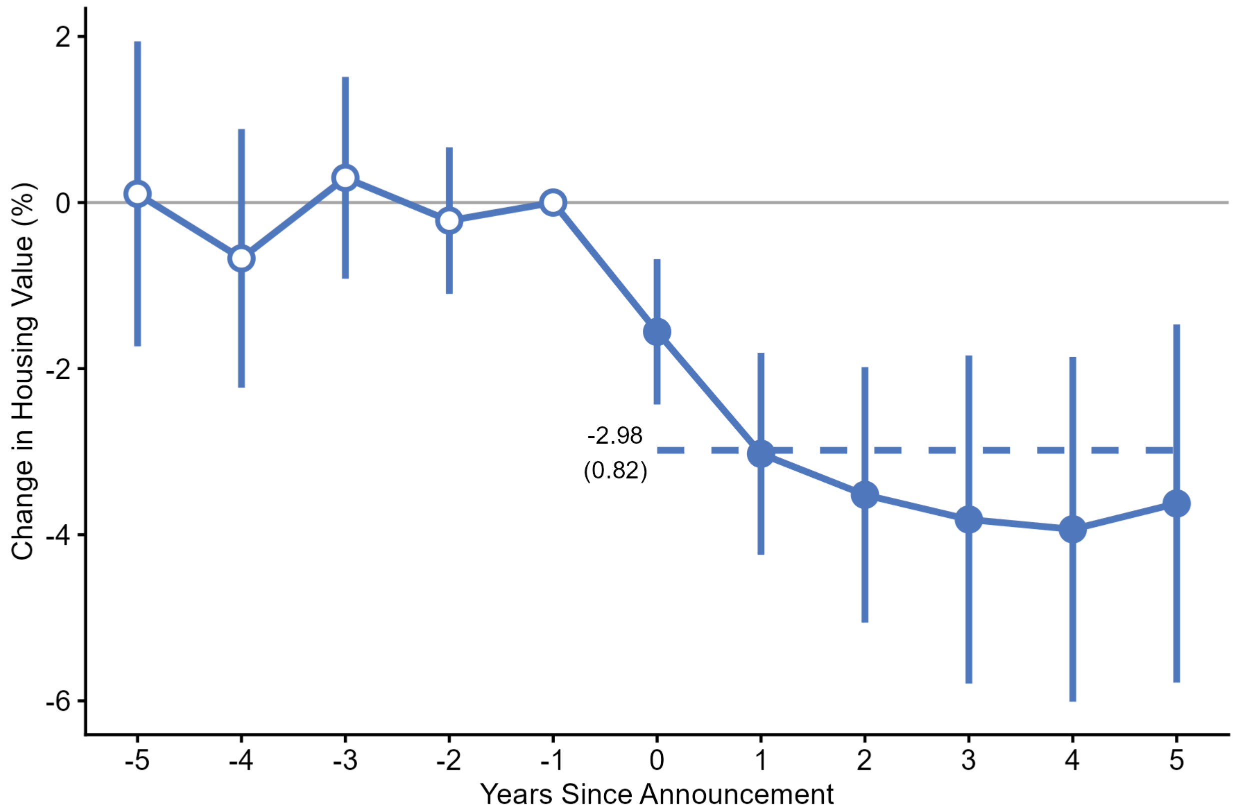
[46] L Sun, S Abraham, Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics* **225**, 175–199 (2021).

[47] K Borusyak, X Jaravel, J Spiess, Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies* **91**, 3253–3285 (2024).

Figure 1: Geographic and Temporal Distribution of Treated and Matched Comparison Data Centers and Housing Transactions

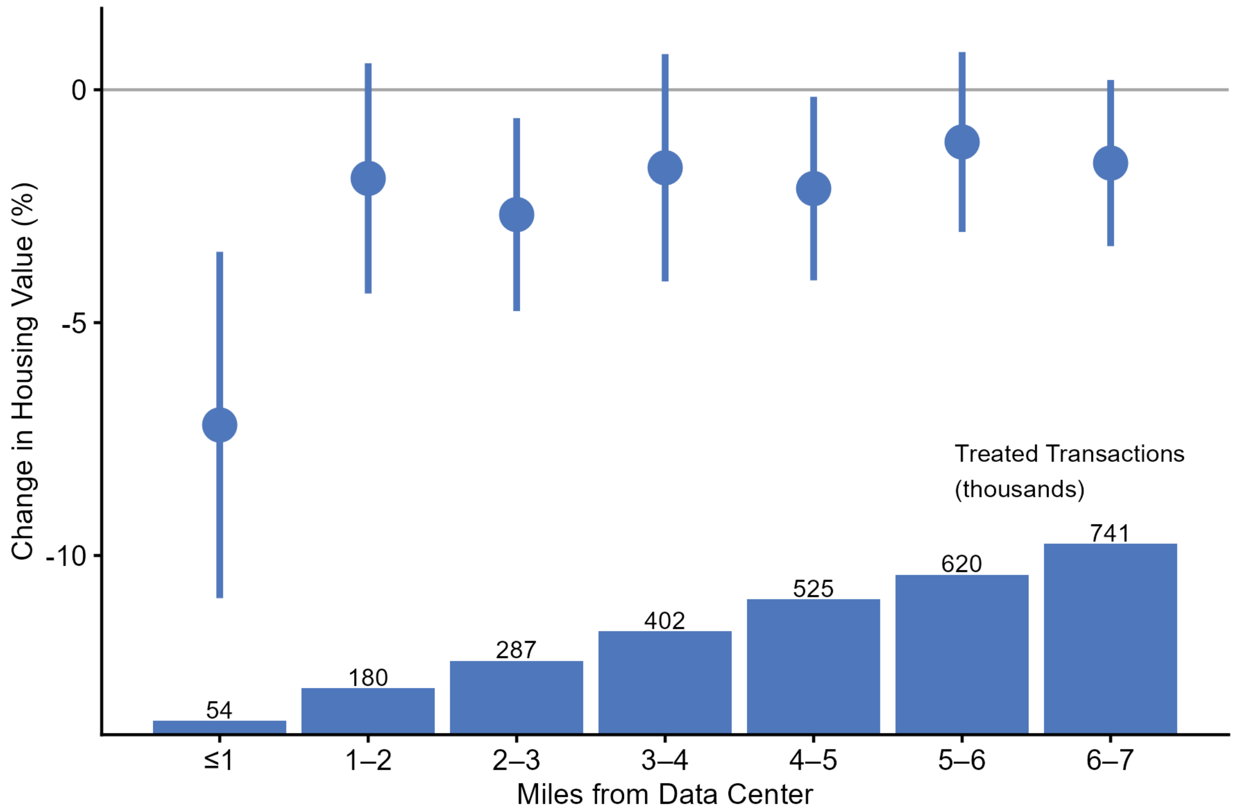
Note: (A) Locations of treated and matched comparison data centers. Filled markers denote treated data centers, colored by announcement year; gray open markers denote matched comparison data centers. Some data centers serve as matched comparisons for earlier-announced data centers and later serve as treatment events themselves. These locations are denoted by both filled and open markers. Marker shape encodes data center type. The large panel shows the contiguous United States; the two stacked panels to the right are zoomed insets for the two Metropolitan Statistical Areas (MSAs) containing the most data centers in our sample: Washington–Arlington–Alexandria ("Washington, DC") and Chicago–Naperville–Elgin ("Chicago, IL"). MSA boundaries are shown in black; state borders are shown in light gray. The black rectangles on the national map mark the extent of each inset. (B) Housing transactions used in the regressions, by treated and matched comparison group. (C) Stacked bar chart showing the distribution of announcement years for treated and matched comparison data centers. (D) Distribution of transaction years for housing transactions in the sample.

Figure 2: Dynamic Effects of Data Centers on Housing Values



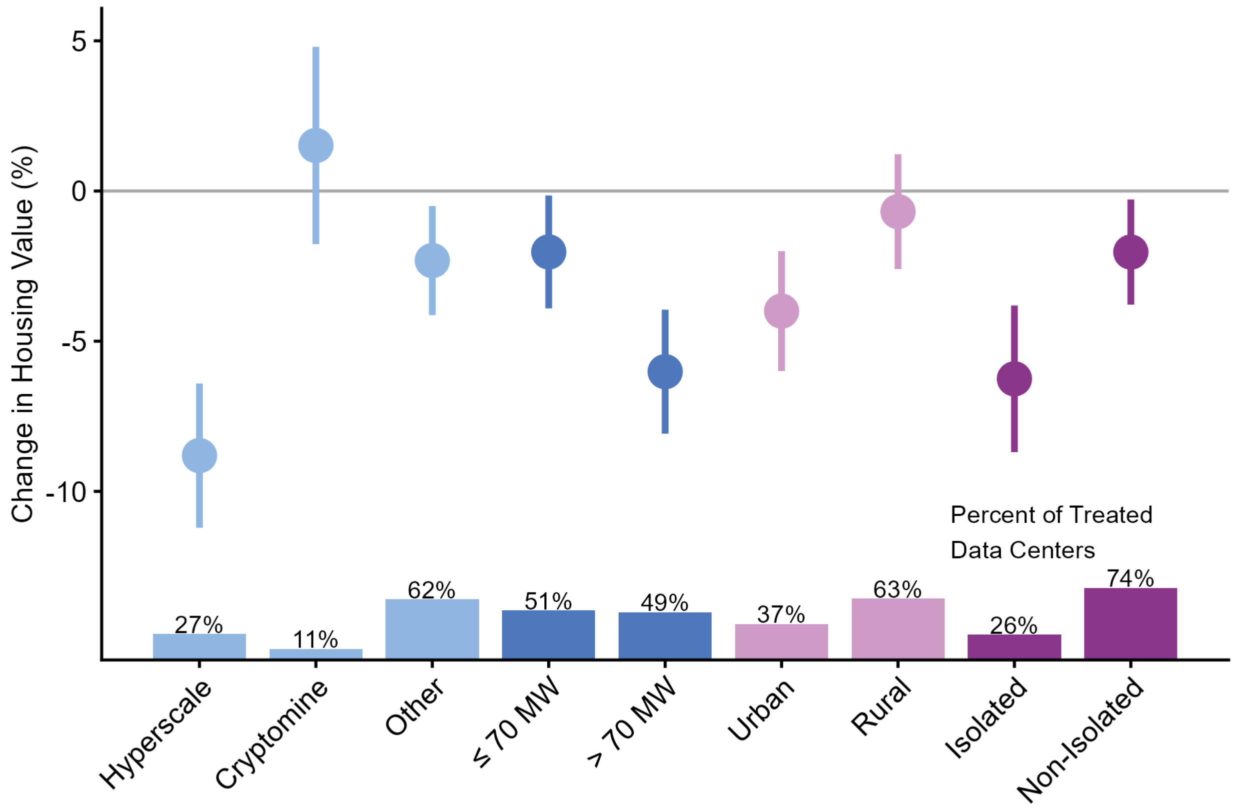
Note: The figure shows estimated changes in housing values before and after the announcement of a large data center relative to the year immediately preceding the announcement, obtained from eqn. (1) using a sample of 2,712,332 observations. Estimates before announcement (hollow dots) provide a test for differential pre-trends in housing values between the treatment and matched comparison groups; estimates after announcement (solid dots) trace how relative housing prices evolve following the announcement. The dashed horizontal line shows the average post-announcement treatment effect estimated by the difference-in-differences regression in eqn. (2). Points denote estimated effects and vertical bars represent 95% confidence intervals based on standard errors clustered at the census tract level. Because the dependent variable is the natural logarithm of the home sale price, coefficients are reported as exact percentage changes, computed as $100(e^{\beta} - 1)$. Standard errors are converted to the percentage scale using the delta method, and the 95% confidence intervals are constructed from these transformed standard errors.

Figure 3: Effect of Data Centers on Housing Values by Distance



Note: The figure shows estimated treatment effects on housing prices by 1-mile distance bins from announced large data centers. Estimates are obtained from eqn. (3) using an extended sample of 4,359,640 observations, which extends the baseline difference-in-differences specification by interacting the $Treat \times Post$ indicators with indicators for each 1-mile distance bin. Points denote the estimated treatment effect for each distance bin relative to the matched comparison group. Vertical bars represent 95% confidence intervals based on standard errors clustered at the census tract level. Because the dependent variable is the natural logarithm of the home sale price, coefficients are reported as exact percentage changes, computed as $100(e^{\beta} - 1)$. Standard errors are converted to the percentage scale using the delta method, and the 95% confidence intervals are constructed from these transformed standard errors. To avoid conflating the distance gradient from the original treatment event with exposure to the announcement or development of a subsequent nearby data center, we exclude transactions for treated homes once another large data center is announced within 7 miles; earlier transactions remain in the sample. Bars at the bottom of the figure show the number of treatment group transactions in each distance bin, in thousands.

Figure 4: Effect of Data Centers on Housing Values by Data Center and Area Characteristics



Note: Circles denote point estimates from eqn. (4) using the main sample of 2,712,332 observations, which augments the baseline difference-in-differences specification by interacting the treatment indicator with mutually exclusive indicators for data center or area characteristics. We examine heterogeneity by data center type (hyperscale, cryptomining, and other), data center size (above versus below the sample median of approximately 70 MW), whether the data center is located in an urban or rural area according to the 2000 Census, and whether the data center is isolated or non-isolated, where isolated data centers are defined as having no other large data center announced or operational within 10 miles before or during their event window. Vertical bars represent 95% confidence intervals based on standard errors clustered at the census tract level. Because the dependent variable is the natural logarithm of the home sale price, coefficients are reported as exact percentage changes, computed as $100(e^{\beta} - 1)$. Standard errors are converted to the percentage scale using the delta method, and the 95% confidence intervals are constructed from these transformed standard errors. Bars at the bottom of the figure show the percentage of treated data centers in each category.

Supporting Information

Data Centers Lower Nearby Housing Values

Garrison Schlauch and Simon Greenhill

Table of Contents

A Data

A.1 Imputing Demographic and Socioeconomic Characteristics

We construct pre-announcement characteristics for the treatment and matched comparison groups by spatially aggregating census block group (CBG)-level data from the 2000 Census and the 2005–2009 through 2018–2022 5-year American Community Surveys (ACSs), obtained from IPUMS NHGIS (1). For each sub-experiment, we aggregate these data over the geographic areas defined by the treatment and comparison group assignment rules described in the Materials and Methods. Treatment group areas consist of the portions within 3 miles of the treated data center assigned to that treatment event; because homes exposed to multiple nearby announcements are assigned only to their first, or in the case of simultaneous announcements, the nearest, data center, these treatment group areas are mutually exclusive and need not form complete 3-mile circles. Comparison group areas, in contrast, are the full 3-mile circles around the matched comparison data centers. This is because candidate comparison data centers are required to be at least 10 miles from any large data center announced or operational by the end of the treated data center's event window (i.e., through five years after its announcement), so the 3-mile area surrounding the comparison data center does not need to be partitioned to account for announcements of other nearby data centers.

We first intersect each treatment and comparison group area with CBG boundaries and let a_{gb} denote the fraction of CBG b's area contained in group g. For count variables X_b (population, the number of housing units, and the number of households), we calculate

$$X_g = \sum_b a_{gb} X_b, \qquad \text{(SI 1)}$$

where $a_{gb} = 0$ for CBGs that do not intersect group g. Thus, each CBG contributes in proportion to the share of its area contained within the treatment or comparison group area. For example, if group g contains 50% of the area of CBG b with 1,000 residents (i.e., $a_{gb} = 0.5$), we impute that 500 of the CBG's residents live within group g.

This procedure implicitly assumes that counts are uniformly distributed within each CBG. Although this assumption is only approximate, CBGs are the smallest geographic units for which these ACS measures are publicly available. The resulting measurement error is likely to be minimal in densely populated areas, where CBGs commonly cover only a few city blocks, but may be greater in sparsely populated areas with geographically larger CBGs. We construct three of the variables used for matching—population density, non-White population share, and housing vacancy rate—from the corresponding area-share-imputed counts.

For median household income, the last variable used for matching, we calculate a household-weighted average of CBG median household incomes:

$$M_g = \frac{\sum_b a_{gb} H_b M_b}{\sum_b a_{gb} H_b}, \qquad \text{(SI 2)}$$

where M_b is median household income and H_b is the number of households in CBG b. Because medians are not additive, M_g approximates, rather than exactly recovers, the median household income of the aggregated treatment or comparison group.

We address missing CBG-level values separately for each variable. We exclude CBGs with missing values and rescale the available data to account for the uncovered portion of the group. We construct a group-level value only when CBGs with non-missing data cover at least 50% of the group's area; otherwise, we set the value to missing. Treatment and matched comparison groups are eligible to be included in the regression samples only if they have non-missing values for every pre-treatment characteristic used in matching: population density, median household income, housing vacancy rate, and the non-White population share.

B Robustness

B.1 Testing for Violations of the Parallel Trends Assumption

We find no statistical evidence against the parallel trends assumption. Visual inspection of the event study estimates (Fig. 2) indicates that the pre-treatment coefficients are approximately flat and centered around zero. We also conduct a joint F-test of the null hypothesis that the pre-treatment event study coefficients are jointly equal to zero and fail to reject the null ($F = 1.89$, p-value $= 0.11$). As discussed in the Materials and Methods, the parallel trends assumption itself is fundamentally untestable because it concerns unobservable counterfactual outcomes. Nevertheless, these results indicate that, conditional on our controls, housing prices in the treatment and comparison groups trended similarly before the treatment event.

B.2 Including Outlier Transactions

Our main specification drops outlier transactions, defined as those below the 1st or above the 99th percentile of either sale price or total room count. These exclusions are intended to reduce the influence of recording errors, highly unusual homes, and other extreme transactions. As a robustness check, we re-estimate the main specification including these transactions.

Table S5, column 2 reports the result of this specification. Including outlier transactions yields an estimated effect of about -0.028 log points (p-value < 0.01), corresponding to a 2.73% decline, which is similar to our main estimate of -2.98%.

B.3 Dropping Housing Characteristic Controls

As a robustness check, we re-estimate our main specification without controlling for the number of rooms and square footage, which we include in the main specification to help account for potential changes in the composition of homes sold over time. The resulting estimate is about -0.033 log points (p-value < 0.01; Table S5, column 3), corresponding to a 3.26% decline, which is similar to our main estimate of -2.98%. This similarity suggests that our results are not driven by changes in the composition of transacted homes as measured by their observable characteristics, but instead reflect changes in the market value of otherwise comparable homes.

B.4 Dropping Observations Treated by Multiple Data Centers

As discussed in the Materials and Methods, our main specification assigns each treated home to the nearest data center first announced within 3 miles. Because data centers often cluster geographically, a home may subsequently be exposed to announcements of additional nearby data centers during the post-treatment event window. Consequently, our baseline estimates may partially capture the effects of these subsequent announcements in addition to the focal announcement.

To assess the importance of this issue, we re-estimate the main specification after dropping transactions once they are exposed to the announcement of another large data center. The resulting estimates can therefore be interpreted more cleanly as reflecting exposure to a single nearby announcement. As expected, the estimated effect of about -0.021 log points (p-value = 0.01; Table S5, column 4), corresponding to a decline of 2.09%, is smaller in magnitude than our baseline estimate of -2.98%, consistent with subsequent nearby announcements contributing to the overall effect. Nevertheless, the estimate remains relatively similar and statistically significant, indicating that our main findings are not primarily driven by exposure to subsequent announcements of other nearby data centers.

B.5 Alternative Standard Errors

In our main specification, we cluster standard errors by census tract rather than treatment or comparison group because the groups in our setting do not correspond to the relevant dependence structure. Homes are treated by the nearest first-announced data center within 3 miles, so geographically adjacent homes may belong to different treatment groups. This can occur when homes are slightly closer to different data centers announced in the same year or when one home lies within 3 miles of an earlier announcement while a nearby home falls just outside that radius and is subsequently assigned to a later announcement. In our sample, 66,335 pairs of homes and 129,229 pairs of transactions assigned to different treatment groups are located within 0.1 miles of one another. Group-level clustering would therefore impose independence across transactions among geographically proximate homes, which may be subject to common local shocks. In addition, overlapping comparison groups can contain transactions among the same not-yet-treated homes, which group-level clustering would also treat as independent. By clustering on geography rather than constructed group membership, census tract clustering allows errors to be correlated among nearby homes even when they are assigned to different groups or appear in different sub-experiments. Table S5, column 5 reproduces our main estimate with more conservative standard errors that two-way cluster by census tract and calendar year. The estimate remains significant at the 10% level (p-value = 0.057).

B.6 Repeat Sales Model

Our findings are robust to using a repeat-sales model. As a robustness check, we replace the group-by-sub-experiment fixed effects, α_{ge}, in our main stacked event study specification (eqn. 1) with home-by-sub-experiment fixed effects, α_{ie}, which absorb all time-invariant characteristics of individual homes. Identification therefore comes from within-home changes in transaction prices over time, comparing treated homes with homes in the corresponding comparison group. Because this approach requires homes to transact multiple times within the event window of a given sub-experiment, it substantially reduces the estimation sample. Nevertheless, we estimate an effect of about -0.048 log points (p-value < 0.01; Table S4, column 2), corresponding to a decline of 4.66%, which is roughly 56% larger in magnitude than our baseline estimate of -2.98%. The corresponding event study continues to exhibit no evidence of differential pre-trends (Fig. S3), and we fail to reject the null hypothesis that the pre-treatment coefficients are jointly equal to zero (F = 1.01, p-value = 0.40). These results indicate that our findings are not driven by changes in the composition of homes sold following the announcement of a large data center and, if anything, are stronger when comparing price changes for the same homes over time.

B.7 Increasing the Number of Matched Comparison Groups

To investigate whether our findings are sensitive to the number of matched comparison groups in each sub-experiment, we match each treated data center to its three closest comparison data centers rather than only the single closest comparison based on the Mahalanobis distance in pre-treatment characteristics. We first select the comparison data center with the smallest Mahalanobis distance, which we use in our main specification. We then exclude all remaining candidate comparison data centers within 10 miles of the selected match, select the remaining candidate with the next-smallest Mahalanobis distance, and repeat this procedure to select the third matched comparison data center. This procedure ensures that matched comparison groups within each sub-experiment are geographically distinct and do not share any homes, while increasing the number of unique matched comparison data centers from 104 in our main specification to 165.

Using a single matched comparison group maximizes the observed similarity between the treated and comparison group areas by selecting the candidate with the smallest Mahalanobis distance. However, it makes the estimated counterfactual more sensitive to idiosyncratic price movements or sparse transaction data in that one comparison group. Using multiple matched comparison groups draws on more locations and transactions, potentially improving precision and reducing dependence on any single comparison group. However, the additional matches may be somewhat less similar to the treatment group.

Table S4, column 3 reports the result of running Eq. (2) using three matched comparison groups in each sub-experiment, with the sub-experiment-by-group fixed effects, α_{ge}, modified so that the treatment group and each matched comparison group within a sub-experiment have their own intercept. The point estimate of about -.026 log points (p-value < 0.01), corresponding to a 2.55% price decline, is similar under this alternative matching procedure to our main estimate of -2.98%. The corresponding event study also exhibits no evidence of differential pre-trends (Fig. S3), and we fail to reject the null hypothesis that the pre-treatment coefficients are jointly equal to zero (F = 0.62, p-value = 0.65). These results indicate that our findings are not sensitive to the number of matched comparison groups used.

B.8 Alternative Weighting Scheme

Wing et al. (2) propose a weighting scheme which, in our context, amounts to weighting each sub-experiment in proportion to the estimated underlying number of treated housing units (see Section A.1), thereby estimating the effect on the average treated housing unit. We do not use this weighting scheme in our main specification because its balanced panel requirement would meaningfully alter the composition of the estimation sample. In particular, it would exclude the 151 treated data centers (43.6%) announced in 2020 or later because they lack a complete post-treatment window. It would also exclude sub-experiments without treatment and comparison group transactions in every event year, disproportionately removing rural data centers, where housing transactions are sparser. The resulting sample of 173 out of the original 346 treated data centers would therefore be concentrated among earlier announcements and more urban areas, limiting the representativeness of the estimates for the broader set of data centers in our study.

We nevertheless implement this weighting scheme as a robustness check. Because our regressions use transaction-level data, we divide each sub-experiment-by-group-by-event year cell weight from (2) by the number of transactions in that cell so that the transaction weights sum to the prescribed cell weight. We estimate this specification using the three closest matched comparison data centers rather than only the closest match because the combination of housing unit weights and, at times, sparse transaction data makes the results more sensitive to idiosyncratic variation in a single comparison group. In sub-experiments with many comparison housing units but few observed transactions, individual comparison transactions receive disproportionately large weights. Using three matches mitigates this issue by spreading the comparison group weights across more transactions.

Table S4, column 4 reports the weighted difference-in-differences estimate. At -0.026 log points (p-value < 0.01), corresponding to a price decline of 2.55%, this effect is similar to our main estimate of -2.98% (column 1). It is also nearly identical to the estimate obtained using three matched comparison groups without weighting (Table S4, column 3). The corresponding event study exhibits no evidence of differential pre-trends (Fig. S3), and we fail to reject the null hypothesis that the pre-treatment coefficients are jointly equal to zero ($F = 2.21$, p-value $= 0.07$).

B.9 Leave-One-Out Estimation

To assess the influence of data from individual states on our main results, we conduct a leave-one-out analysis in which we repeatedly re-estimate the main difference-in-differences specification (eqn. 2), each time excluding all sub-experiments whose treated data center is located in a given state. Fig. S4 presents the resulting point estimates and 95% confidence intervals. Across all 32 leave-one-out regressions, the estimated effects remain statistically significant at the 5% level, and every estimate falls within the 95% confidence interval of the main specification.

C Additional Analyses

C.1 Balance Between Treatment and Comparison Groups

The treatment and matched comparison groups are relatively well-balanced on the pre-treatment characteristics used for matching. To assess balance, we compute standardized mean differences (SMDs) for each matching characteristic. When 5-year ACS estimates are available, we define each characteristic's pre-treatment value as the average over event years $t = -5$ through $t = -1$, using the corresponding 5-year ACS estimates. Data centers announced from 2010 through 2023 therefore use ACS vintages ranging from 2006–2009 through 2018–2022, respectively. For data centers announced before 2010, for which corresponding pre-treatment ACS estimates are unavailable, we use the 2000 Census.

Let E be the number of sub-experiments, each containing one treated group and one matched comparison group. For a matching characteristic X, define the within-sub-experiment difference and its mean as

$$d_e = X_{eT} - X_{eC}, \tag{SI 3}$$

$$\bar{d} = \frac{1}{E}\sum_{e=1}^{E} d_e = \bar{X}_T - \bar{X}_C, \tag{SI 4}$$

where X_{eT} and X_{eC} are the characteristic values for the treated and matched comparison group in sub-experiment e, respectively.

We measure the magnitude and direction of imbalance in X using the standardized mean difference:

$$\text{SMD} = \frac{\bar{d}}{s_{\text{pool}}}, \qquad s_{\text{pool}} = \sqrt{\frac{s_T^2 + s_C^2}{2}}, \tag{SI 5}$$

where s_T^2 and s_C^2 are the variances of the characteristic across the treatment and matched comparison groups, respectively. An SMD of 1, for example, indicates that the treatment group mean is one pooled standard deviation higher than the matched comparison group mean.

Table S3 reports the SMDs and corresponding treatment and comparison group means. The SMDs range from 0.05 to 0.21, all below the commonly used threshold of 0.25 (3). The remaining imbalance is consistent with evidence that planned data centers have increasingly been sited in more rural areas over time (4). Accordingly, the treated data centers in our sample, which were announced an average of 7 years earlier than the matched comparison data centers (Table S2), are located in areas with somewhat higher population density (SMD = 0.21), median household income (SMD = 0.15), and the percentage of the population that is non-White (SMD = 0.13).

C.2 Announcement Versus Opening Effects

We find that the timing of housing-price capitalization changes markedly over the sample period. Among data centers announced from 2003 through 2020, we find no evidence that housing prices respond at announcement but estimate a statistically significant additional decline once the data center becomes operational. In contrast, among data centers announced from 2021 through 2023—the beginning of the post-pandemic hyperscale and AI boom (Fig. S1), which coincides with an upward trend break in Google Search interest in data centers (Fig. S5)—housing prices decline at announcement, and we detect no additional effect after opening. This shift is consistent with the idea that, as data centers have become more salient in recent years and their potential local impacts are better understood, housing markets increasingly capitalize their expected effects when data centers are announced rather than waiting until they become operational.

To disentangle housing price responses to the announcement of a large data center from those associated with the opening, we impose three sample restrictions. First, we restrict the sample to the 237 out of 346 treated data centers from our main sample that became operational by 2024, the end of our sample period. Second, we require that the treated data center be announced and become operational in different calendar years, allowing us to separately identify the announcement and opening effects. This restriction further reduces the sample to 218 sub-experiments. Finally, we require that the treated data center become operational within 4 calendar years of announcement, ensuring that the opening occurs within the 5-year event window and that at least one subsequent event year is observed. This final restriction leaves 215 treated data centers: 181 announced before 2021 and 34 announced after 2021.

Within this sample, we distinguish capitalization at announcement from any additional effect associated with operation by estimating a specification that extends our main difference-in-differences model (eqn. 2) to include separate indicators for the post-announcement and post-opening periods. We estimate this equation separately for the two sets of data centers described above—those announced from 2003–2020 and those announced from 2021–2023:

$$\begin{aligned}\log(price)_{ige\tau t} =& \beta_A \left(PostAnnouncement_{e\tau} \times Treat_{ige}\right) \\ &+ \beta_O \left(PostOpening_{e\tau} \times Treat_{ige}\right) \\ &+ \alpha_{ge} + \delta_{e\tau} + \mathbf{X}'_{ige\tau t}\gamma + \varepsilon_{ige\tau t}.\end{aligned} \tag{SI 6}$$

Here, $PostAnnouncement_{e\tau}$ is an indicator equal to one in the announcement year and all subsequent years (i.e., the same as $Post_{e\tau}$ in eqn. 2), while $PostOpening_{e\tau}$ is an indicator equal to one beginning in the opening year. Because $PostAnnouncement_{e\tau}$ remains equal to one after the data center becomes operational, β_A captures the average differential change in housing prices after announcement but before opening. The coefficient β_O captures the incremental change beginning in the opening year relative to the announcement-to-opening period.

For the 181 data centers announced from 2003 through 2020 that meet the above criteria, we find no evidence that housing prices responded at announcement, and almost the entire price effect is observed after opening. We fail to reject the null hypothesis that $\beta_A = 0$ at conventional significance levels, and the point estimate of about -0.006 log points, corresponding to a 0.60% decline, is relatively small in magnitude (Table S6A, column 2). In contrast, we reject the null hypothesis that $\beta_O = 0$ at the 1% level (Table S6A, column 2). Moreover, the incremental post-opening effect of -0.030 log points (-2.97%) is similar to the overall post-announcement effect of -0.026 log points (-2.58%) estimated using our main difference-in-differences specification (eqn.2) for this sample (Table S6A, column 1). Together, these estimates suggest that, for data centers announced through 2020, nearby housing prices responded minimally at announcement and declined almost entirely after data centers became operational.

For the 34 data centers meeting the above criteria that were announced from 2021–2023, we find the opposite pattern. We reject the null hypothesis that $\beta_A = 0$ at the 5% level (Table S6B, column 2), and the estimated announcement effect is about -0.041 log points (-3.97%), which is nearly identical to the overall post-announcement effect of roughly -0.041 log points (-3.98%) from our main difference-in-differences specification for this sample (Table S6B, column 1). In contrast, we fail to reject the null hypothesis that $\beta_O = 0$ at conventional significance levels, and the estimated incremental opening effect of -0.01% is approximately zero (Table S6B, column 2). These estimates suggest that, for more recently announced data centers, nearby housing prices adjusted when data centers were announced, with no detectable additional effects once data centers became operational.

Lastly, the announcement-to-opening interval is similar across the two periods, averaging 1.88 years for the 181 data centers announced from 2003–2020 and 1.59 years for the 34 announced from 2021–2023. Thus, the shift toward capitalization at announcement in the later period does not appear to be driven by systematically shorter lags between announcement and opening, which could otherwise leave less time to detect separate price responses.

C.3 Sales Impacts

We find no clear evidence that the announcement of a large data center affects housing sales, suggesting that our estimated price effects are unlikely to be driven by large changes in transaction volume. To examine whether data center announcements affect transaction volume or the probability of observing a transaction, we aggregate the data to the sub-experiment-by-group-by-census tract-by-event year level. Let $N_{gce\tau}$ denote the number of housing sales in treatment or matched comparison group g, census tract c, and event year τ of sub-experiment e. We construct a balanced panel over all observable cells, including every event year that can be observed within our sample period. For cells with no observed sales, we set $N_{gce\tau} = 0$. Approximately 12.5% of cells contain no sales.

Using these grouped data, we estimate a stacked event study specification analogous to our main specification (eqn. 1):

$$Y_{gce\tau} = \sum_{h=-5,\ldots,5 \;\; h\neq -1} \beta_h \left(\mathbf{1}\{\tau = h\} \times Treat_{gce}\right) + \alpha_{ge} + \delta_{e\tau} + \varepsilon_{gce\tau}, \qquad \text{(SI 7)}$$

We estimate Eq. (SI 7) using two dependent variables. First, we set $Y_{gce\tau} = \log(N_{gce\tau} + 1)$, which measures changes in the number of home sales while retaining cells with no sales. Second, we set $Y_{gce\tau} = \mathbf{1}\{N_{gce\tau} > 0\}$, which indicates whether any sale is observed in the cell. The first outcome captures changes in transaction volume, while the second tests whether announcements affect the probability of observing any sale. As in our main specification, we cluster the standard errors at the census tract level.

We also estimate the corresponding difference-in-differences specification using each of the two dependent variables described above:

$$Y_{gce\tau} = \beta\,(Treat_{gce} \times Post_{\tau}) + \alpha_{ge} + \delta_{e\tau} + \varepsilon_{gce\tau}. \tag{SI 8}$$

Fig. S6 reports the results from Eqs. (SI 7) and (SI 8). The event study coefficients provide no clear evidence of a systematic change in either transaction volume or the probability that any sale occurs following an announcement. Consistent with this pattern, the difference-in-differences estimate implies a 0.98% decline in sales, but the estimate is statistically insignificant (p-value $= 0.52$). The estimated effect on the probability of observing at least one sale in a cell is approximately zero and is also statistically insignificant (p-value $= 0.80$). The event study and difference-in-differences estimates are fairly noisy, however, and therefore do not rule out modest changes in transaction activity.

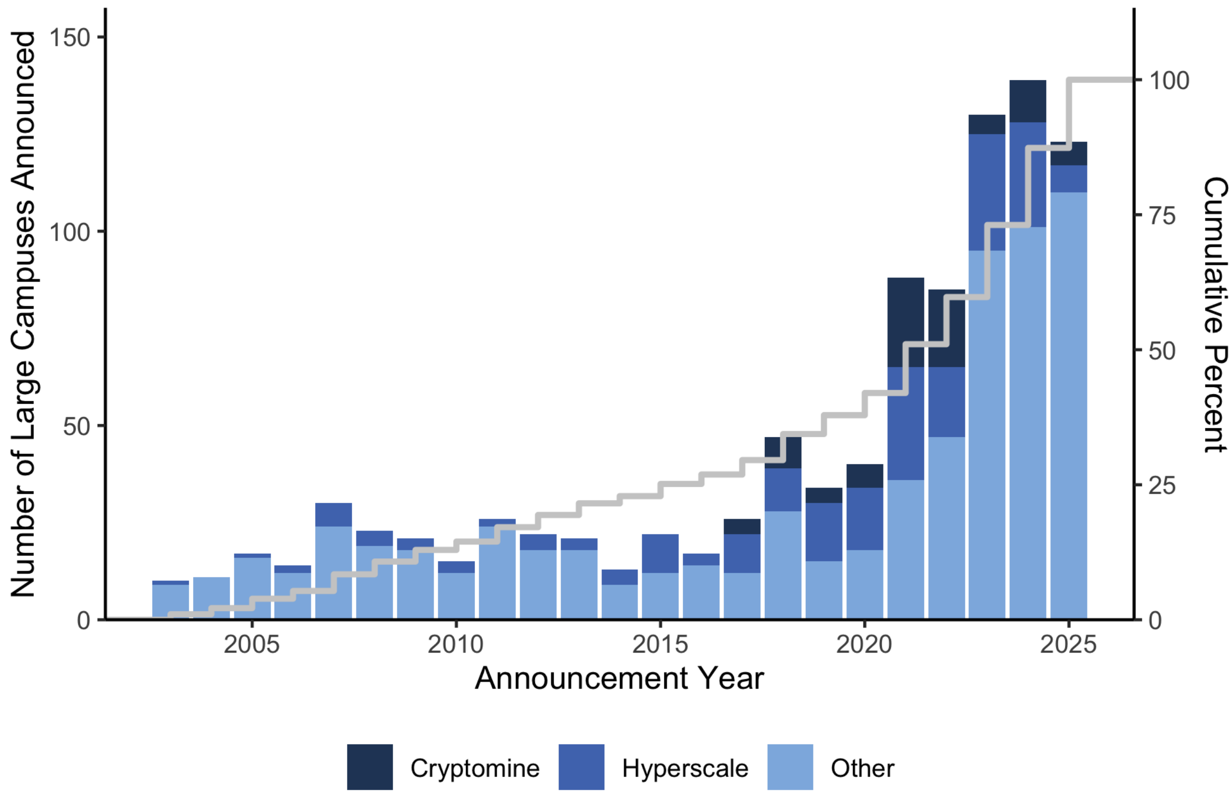
C.4 Noise and Air Quality Monitor Coverage

We investigated whether public environmental monitors could be used to estimate the effect of data center openings on local air quality and noise levels. For air quality, we focused on NO_2 and $PM_{2.5}$ monitors in the EPA's Air Quality System (AQS) network, which measure some of the primary pollutants produced by diesel exhaust from on-site backup generators. For noise, we used the EHZ and HHZ channels of surface-level seismometers, which have been shown to capture anthropogenic noise (5). We consider a monitor potentially usable if it is within 6 miles of a large data center, reports at hourly or finer time resolution, and has 50% or better temporal coverage in each year before and after a data center's opening.

We find that direct measurement of data center noise and local air quality impacts is infeasible at scale with existing public monitoring infrastructure. Fig. S7 shows that two-thirds of large data centers have no qualifying monitor of any kind within 6 miles, and none has one within half a mile. The lack of monitoring near data centers may be especially important for noise, as it attenuates more quickly over distance.

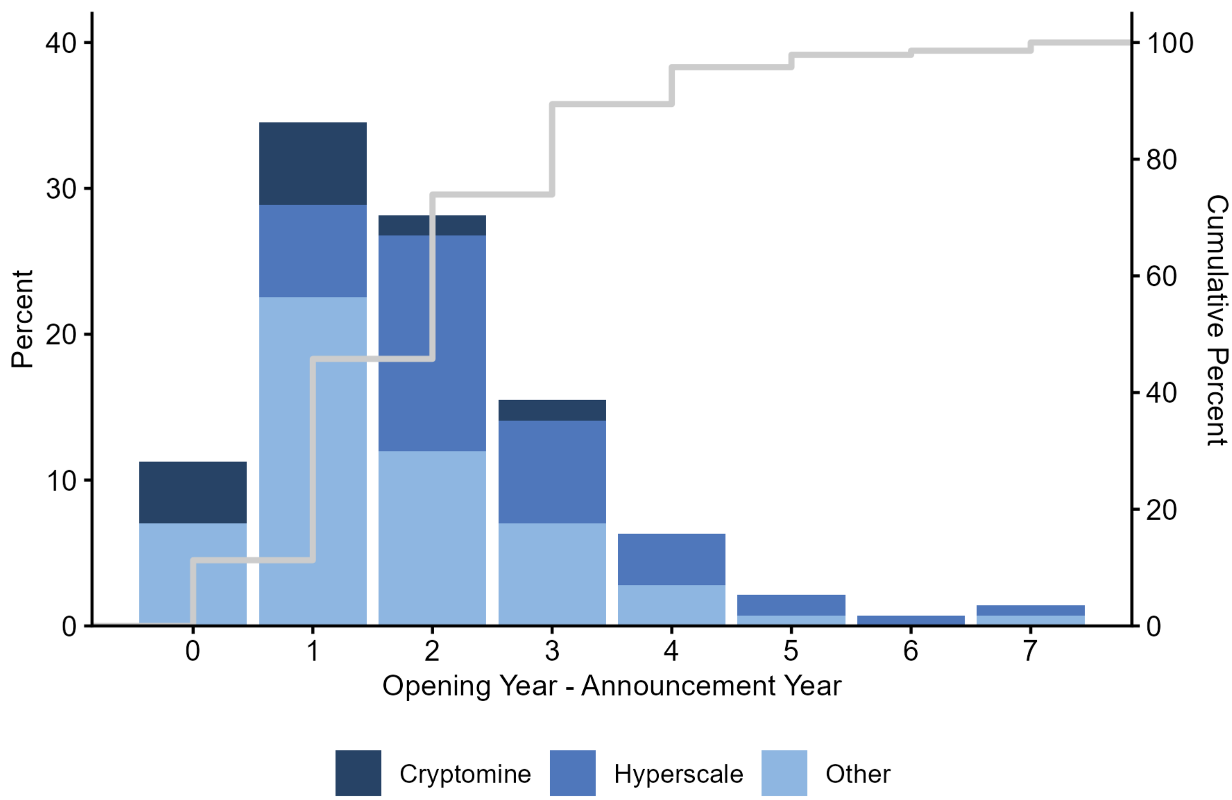
D Supporting Figures

Figure S1: Cumulative Distribution of Data Center Announcement Years



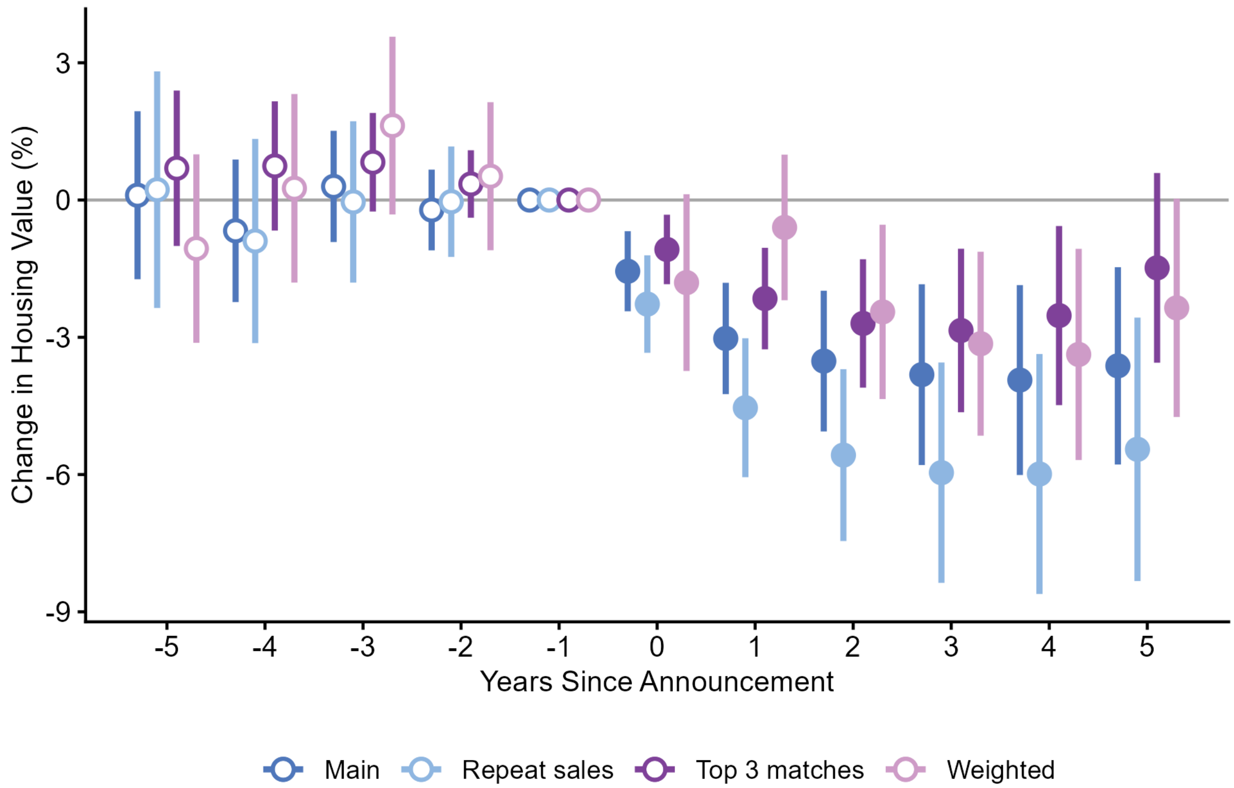
Note: Bars show the number of large data center campuses announced each year between 2003 and 2025, colored by type. The gray line shows the cumulative percentage. Large campuses are defined as having a total power capacity of at least 20 MW or a building area of at least 100,000 square feet.

Figure S2: Distribution of Opening Versus Announcement Years



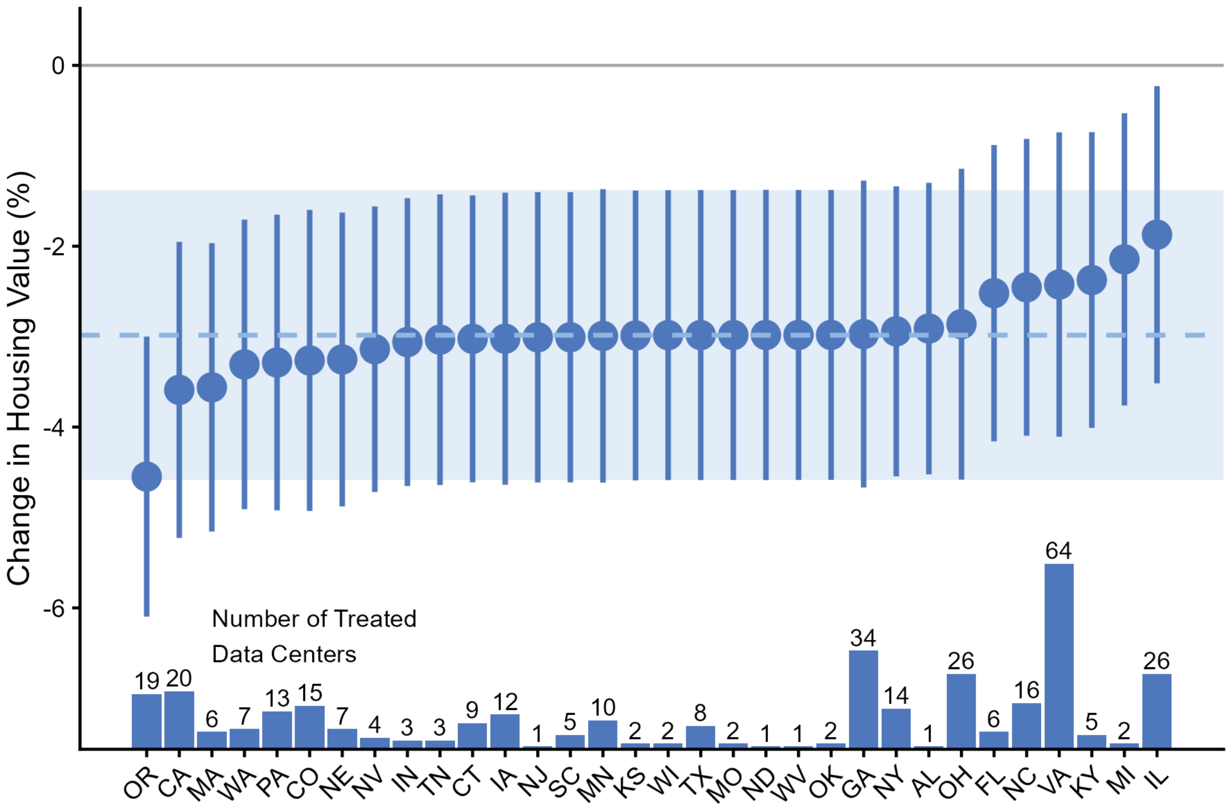
Note: The figure shows the distribution of treated data centers by the number of calendar years between announcement and opening, separately by type: cryptomining (dark blue), hyperscale (mid-blue), and other (light blue). The sample includes the 142 out of 346 treated data centers that (i) were operational by the end of 2025 and (ii) have an observed announcement year that was not imputed using either the construction start year or the median gap between announcement and opening among data centers of the same type. The horizontal axis displays the number of calendar years between opening and announcement. A value of zero indicates that a campus was announced in the same calendar year that it became operational; positive values indicate that it opened in a calendar year after it was announced. The stacked bars report the percentage of data centers at each value on the left vertical axis. The gray line reports the cumulative percentage of data centers on the right vertical axis. Approximately 46% of data centers opened within 1 calendar year of their announcement, 74% within 2 calendar years, and 90% within 3 calendar years. Among data centers with observed announcement years, the median gap between announcement and opening is 1 year for cryptomining and 2 years for hyperscale and other types of data centers.

Figure S3: Robustness to Alternative Research Designs: Event Studies



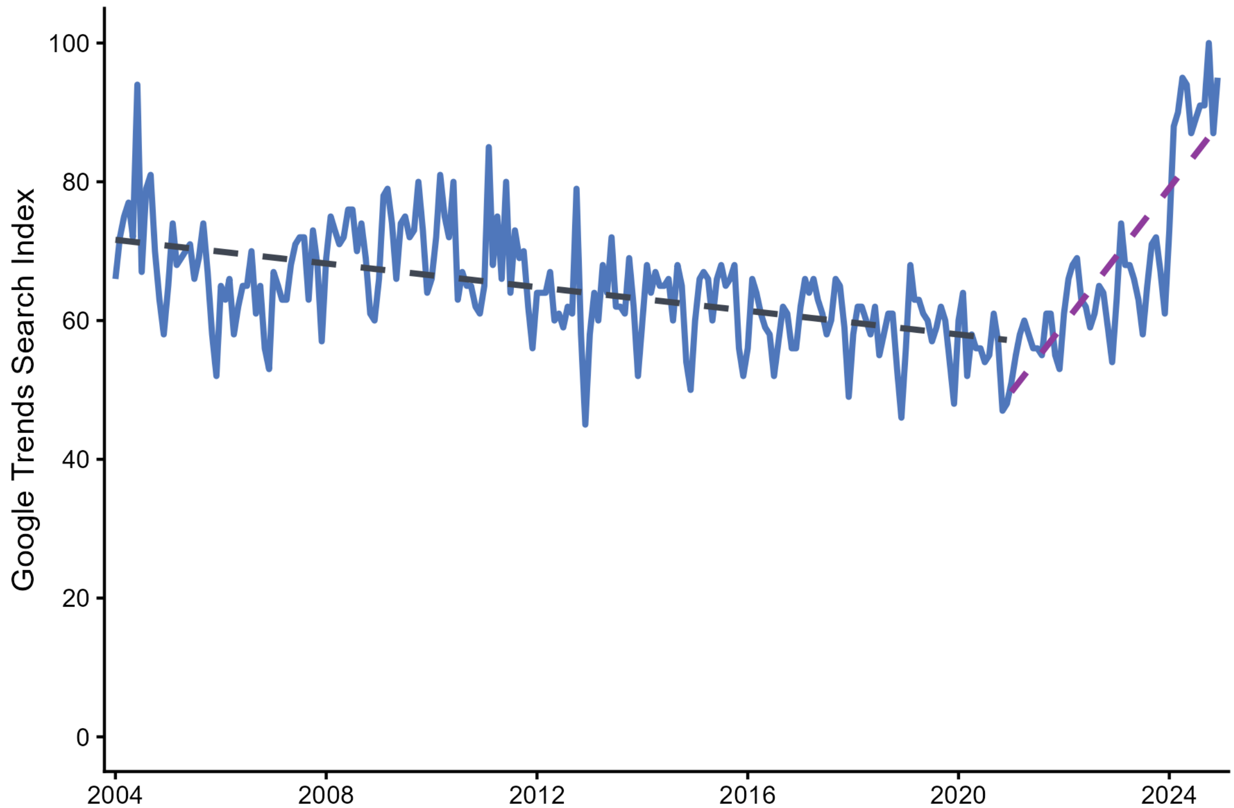
Note: The figure shows estimated changes in housing values before and after the announcement of a large data center relative to the year immediately preceding the announcement, obtained from eqn. (1) under four alternative specifications. The mid-blue estimates correspond to the main specification. The light-blue estimates use a repeat-sales specification that replaces the sub-experiment-by-group fixed effects with sub-experiment-by-home fixed effects, restricting the sample to homes that transact multiple times within a given sub-experiment. The mid-purple estimates use the top three best-matched comparison groups for each treatment event, rather than only the single best match used in the main specification. The light-purple estimates weight sub-experiments by their estimated number of treated housing units using weights derived from those in Wing et al. (2). Estimates before announcement (hollow dots) provide a test for differential pre-trends in housing values between the treatment and comparison groups; estimates after announcement (solid dots) trace how relative housing prices evolve following the announcement. Points denote estimated effects, and vertical bars represent 95% confidence intervals based on standard errors clustered at the census tract level. Because the dependent variable is the natural logarithm of the home sale price, coefficients are reported as exact percentage changes, computed as $100(e^{\beta} - 1)$. Standard errors are converted to the percentage scale using the delta method, and the 95% confidence intervals are constructed from these transformed standard errors.

Figure S4: Leave-One-Out Estimates



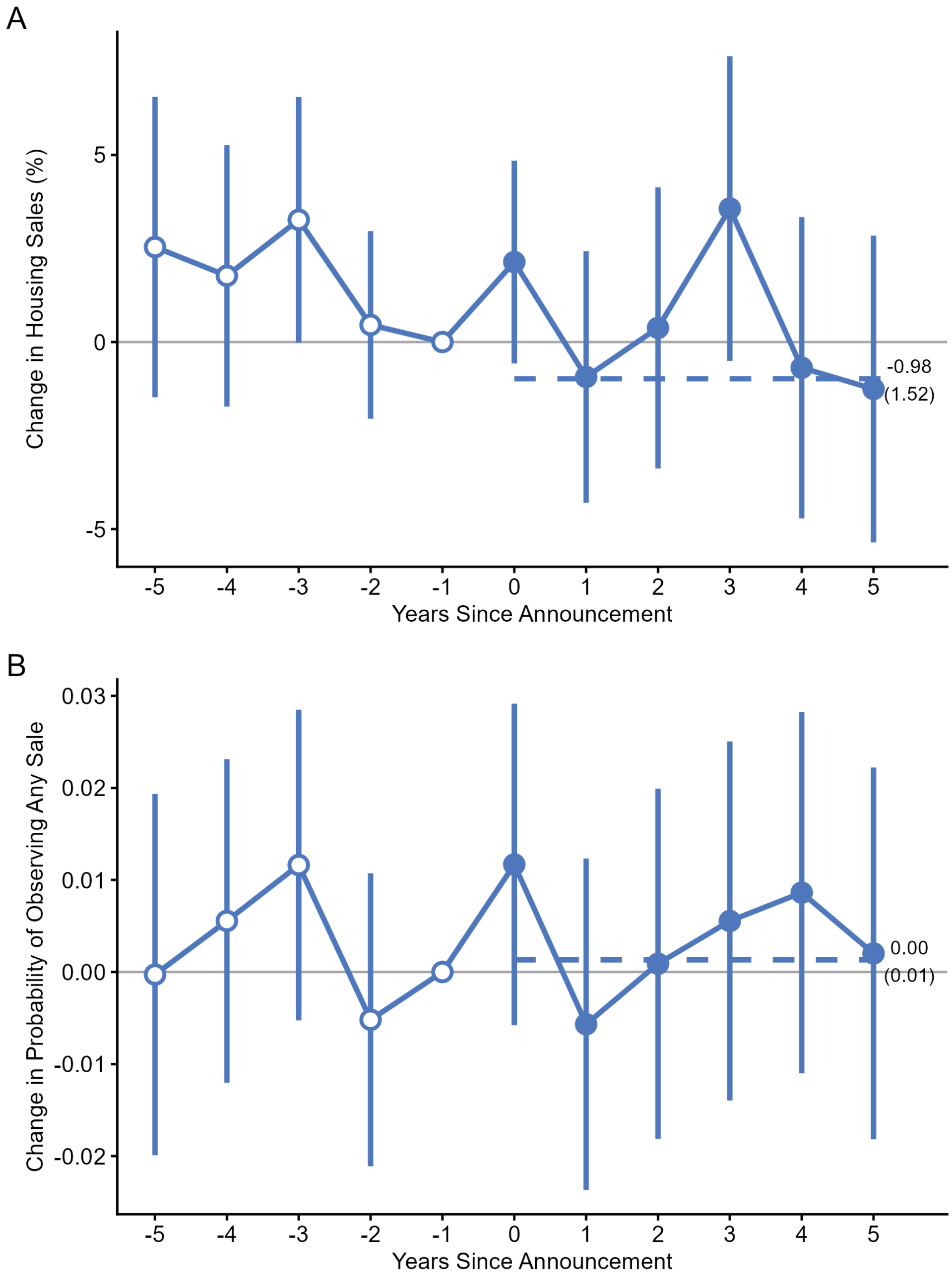
Note: The figure shows the distribution of leave-one-out difference-in-differences estimates by state. Estimates are obtained from 32 separate regressions estimated using eqn. 2, with each regression excluding all sub-experiments whose treated data center is located in a given state indicated on the x-axis. Points denote the estimated treatment effects, and vertical bars represent 95% confidence intervals based on standard errors clustered at the census tract level. The dashed line shows the point estimate from the main specification, and the shaded region shows the corresponding 95% confidence interval. Because the dependent variable is the natural logarithm of the home sale price, coefficients are reported as exact percentage changes, computed as $100(e^{\beta}-1)$. Standard errors are converted to the percentage scale using the delta method, and the 95% confidence intervals are constructed from these transformed standard errors. The bars at the bottom of the figure show the number of treated data centers in each state.

Figure S5: Google Search Interest in Data Centers, 2004–2024



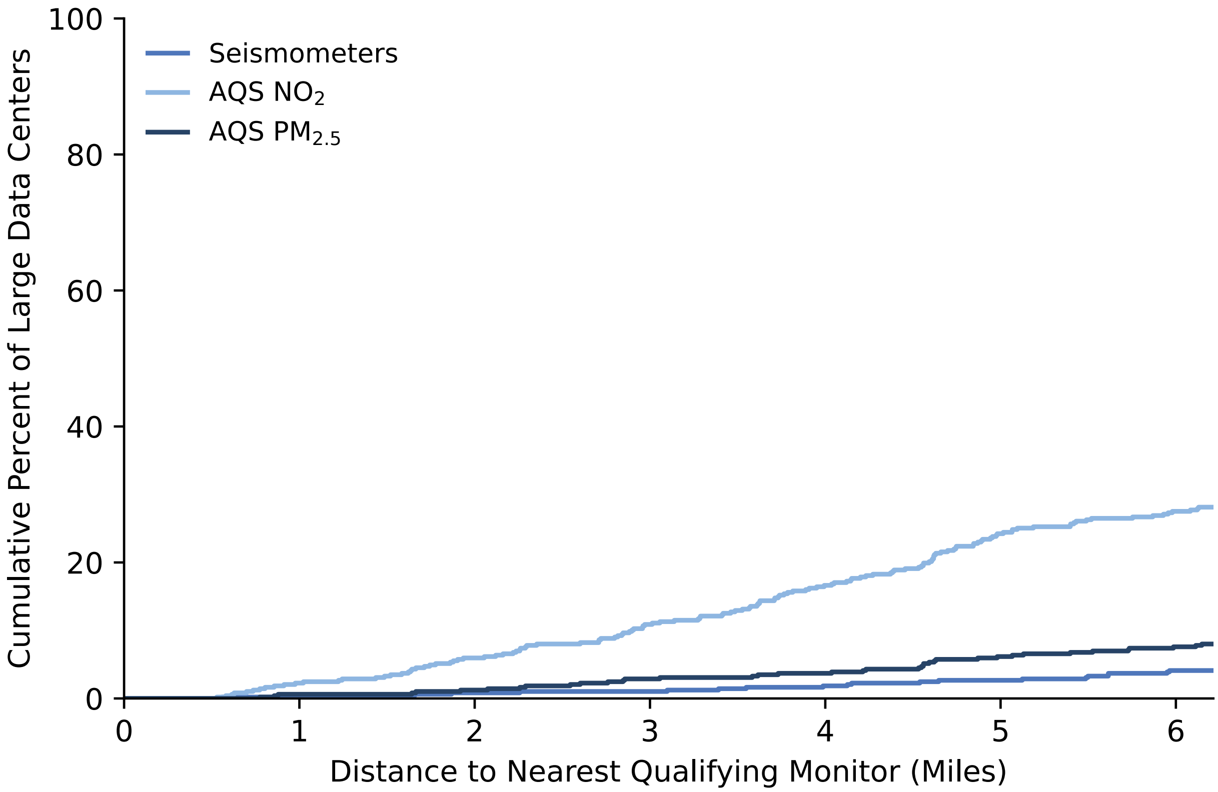
Note: This figure plots monthly U.S. Google Trends search interest for the term "data center" from 2004, the earliest year available, through 2024, the last year in our sample period. The search index is normalized from 0 to 100, with 100 representing the month with the highest search interest over the period. The dashed lines show separate linear trends estimated before (2004–2020) and during (2021–2024) the post-pandemic hyperscale and AI boom.

Figure S6: Effect of Data Centers on Housing Sales



Note: Panel A uses the natural logarithm of the number of home sales in a sub-experiment-by-group-by-census tract-by-event year cell as the dependent variable. Panel B uses an indicator equal to one if the cell contains at least one home sale. Both panels report event study and difference-in-differences estimates from Eqs. (SI 7) and (SI 8). Hollow dots denote pre-announcement event study estimates, which provide a test for differential pre-trends between the treatment and comparison groups, while solid dots denote post-announcement estimates. The dashed horizontal line indicates the average post-announcement treatment effect from the difference-in-differences specification, and the text within each panel reports the corresponding estimate and standard error. Vertical bars represent 95% confidence intervals. In panel A, coefficients are reported as exact percentage changes, computed as $100(e^{\beta}-1)$. Standard errors and confidence intervals are converted to the percentage scale using the delta method. Standard errors in both the event study and difference-in-differences specifications are clustered at the census tract level.

Figure S7: AQS and Seismometer Coverage near Data Centers



Note: The figure shows the cumulative percentage of large data centers with at least one usable environmental monitor within a given distance for three monitor types: surface seismometers, which can measure noise; EPA AQS NO_2 monitors; and EPA AQS $PM_{2.5}$ monitors. A monitor is considered usable if it is within 6 miles of a data center, reports at hourly or finer time resolution, and has 50% or better temporal coverage in each year before and after the data center's opening.

E Supporting Tables

Table S1: Summary Statistics of Housing Transactions

	All	Treatment	Matched comparison
Sale price ($2025)	372,625	386,841	336,931
	(241,396)	(247,679)	(218,938)
Living area (sq. ft.)	1,886	1,889	1,878
	(964)	(1,003)	(845)
# Rooms	5.7	5.7	5.8
	(1.6)	(1.6)	(1.6)
# Bedrooms	3.2	3.2	3.3
	(0.9)	(0.9)	(0.9)
# Bathrooms	2.5	2.5	2.5
	(1.0)	(0.9)	(1.0)
Homes	915,453	721,968	291,560
Transactions	1,366,536	968,603	464,193

Note: The table reports summary statistics for housing transactions in the main sample. Transactions are observed from January 1998 through May 2024 and, within each sub-experiment, are restricted to the same $\pm$5-year window around the treated data center's announcement for both the treatment and comparison groups. The treatment group consists of homes within 3 miles of their first announced large data center, while the comparison group consists of homes within 3 miles of the matched future large data center. Reported variables are inflation-adjusted sale price (2025 dollars), living area square footage, number of rooms, number of bedrooms, the number of bathrooms. All statistics are computed using unique transactions. In the regression sample, however, some transactions appear in multiple sub-experiments because the same comparison group may be used for multiple treatment events, different comparison groups may overlap, or a home's earlier transactions may serve as comparison observations before the home is later treated.

Table S2: Summary Statistics of Data Centers

	All	Treatment	Matched comparison
Panel A: Means and standard deviations			
Announcement year	2018	2016	2023
	(7)	(6)	(3)
Power capacity (MW)	170	122	318
	(309)	(156)	(531)
Square footage (thousands)	934	767	1,451
	(1,280)	(1,083)	(1,651)
Panel B: Percentages			
Urban	34%	37%	22%
Rural	66%	63%	78%
Operational	61%	72%	20%
Planned	39%	28%	80%
Hyperscale	24%	27%	18%
Cryptomining	10%	11%	9%
Other	66%	62%	73%
Number of data centers	415	346	104

Note: The table reports summary statistics for the characteristics of data centers in the main sample. Column (1) reports statistics for all data centers, Column (2) for data centers used as treatment events, and Column (3) for data centers used as matched comparisons. Panel A reports the mean announcement year, MW, and square footage, with standard deviations in parentheses. Panel B reports the percentage of data centers by location (urban or rural), operational status (operational or planned), and data center type (hyperscale, cryptomining, or other). Some matched comparison data centers later serve as treatment events in subsequent sub-experiments and therefore also appear in the treatment sample.

Table S3: Balance of Baseline Characteristics

	Treatment mean	Matched comparison mean	Standardized mean difference
Population per square mile	1,852	1,403	0.21
Median household income ($2025)	104,526	98,382	0.15
Housing vacancy rate (percent)	7.7	7.4	0.05
Non-white population (percent)	35.8	32.8	0.13

Note: The table reports standardized mean differences for the pre-treatment characteristics used to match treated data centers to comparison data centers. Pre-treatment characteristics are defined using the average over event years $t = -5$ through $t = -1$ from the corresponding 5-year ACS estimates when available; for data centers announced before 2010, we instead use the 2000 Census. Standardized mean differences are calculated as the average within-sub-experiment difference between the treatment and matched comparison groups, divided by the pooled standard deviation. Positive values indicate that the treatment group mean is higher than the matched comparison group mean.

Table S4: Robustness to Alternative Research Designs: Difference-in-Differences

	Main (1)	Repeat sales (2)	Top 3 matches (3)	Weighted (4)
Treat × Post	-0.030***	-0.048***	-0.026***	-0.026***
	(0.008)	(0.011)	(0.008)	(0.008)
Fixed effects:				
Sub-experiment × group	✓		✓	✓
Sub-experiment × home		✓		
Sub-experiment × event year	✓	✓	✓	✓
Housing controls	✓		✓	✓
Outcome mean	12.545	12.512	12.545	12.538
Observations	2,712,332	1,252,143	5,147,202	3,720,140
Treated data centers	346	337	346	173
Comparison data centers	104	97	165	134

Note: The table reports difference-in-differences estimates of the change in housing values following the announcement of a large data center, obtained from eqn. (2) under four alternative specifications. The dependent variable is the natural logarithm of the home sale price. Column (1) reports estimates from the main specification. Column (2) uses a repeat-sales specification that replaces the sub-experiment-by-group fixed effects with sub-experiment-by-home fixed effects, which restricts the sample to homes that transact multiple times within a given sub-experiment and omits housing controls since they are collinear with these fixed effects. Column (3) uses the top three best-matched comparison groups for each treatment event, rather than only the single best match used in the main specification. Column (4) applies the weighting approach of (2) to our setting, weighting sub-experiments in proportion to the estimated number of treated housing units; see Section B.8 for more details. Standard errors are clustered at the census tract level and are reported in parentheses. Asterisks denote p-values: * < 0.10, ** < 0.05, *** < 0.01.

Table S5: Robustness to Alternative Specifications: Difference-in-Differences

	Main (1)	Keep outliers (2)	Drop hous. controls (3)	Drop multi-treated (4)	Two-way cluster (5)
Treat × Post	-0.030***	-0.028***	-0.033***	-0.021**	-0.030*
	(0.008)	(0.009)	(0.009)	(0.008)	(0.015)
Fixed effects:					
Sub-experiment × group	✓	✓	✓	✓	✓
Sub-experiment × event year	✓	✓	✓	✓	✓
Housing controls	✓	✓		✓	✓
Standard error clustering:					
Census tract	✓	✓	✓	✓	✓
Calendar year					✓
Outcome mean	12.545	12.524	12.545	12.542	12.545
Observations	2,712,332	2,784,051	2,712,332	2,671,245	2,712,332
Treated data centers	346	346	346	346	346
Comparison data centers	104	104	104	104	104

Note: The table reports difference-in-differences estimates of the change in housing values following the announcement of a large data center, obtained from eqn. (2) under five alternative specifications. The dependent variable is the natural logarithm of the home sale price. Column (1) reports estimates from the main specification. Column (2) retains outliers that are excluded from the main specification. Column (3) omits controls for housing characteristics. Column (4) drops transactions occurring after a home becomes exposed to the announcement of multiple large data centers within 3 miles. Column (5) uses the main specification but two-way clusters the standard errors by census tract and calendar year. In all other specifications, standard errors are clustered at the census tract level and are reported in parentheses. Asterisks denote p-values: * < 0.10, ** < 0.05, *** < 0.01.

Table S6: Data Center Announcement Versus Opening Effects

	(1)	(2)
Panel A: Before the post-pandemic hyperscale and AI boom		
Treat × Post Announcement	-0.026***	-0.006
	(0.010)	(0.010)
Treat × Post Opening		-0.030***
		(0.009)
Outcome mean	12.530	12.530
Observations	1,818,992	1,818,992
Treated data centers	181	181
Comparison data centers	82	82
Panel B: During the post-pandemic hyperscale and AI boom		
Treat × Post Announcement	-0.041***	-0.041**
	(0.014)	(0.016)
Treat × Post Opening		-0.0001
		(0.012)
Outcome mean	12.574	12.574
Observations	186,094	186,094
Treated data centers	34	34
Comparison data centers	19	19
Fixed effects:		
Sub-experiment × group	✓	✓
Sub-experiment × event year	✓	✓
Housing controls	✓	✓

Note: The table compares the estimated effects of the announcement versus opening of a large data center on local housing prices across earlier and later portions of the sample period. The dependent variable is the natural logarithm of the home sale price. Panel A includes data centers announced from 2003–2020, while panel B includes data centers announced from 2021–2023. In both panels, the sample is restricted to sub-experiments in which the treated data center (i) became operational in 2024 or earlier, so that we observe housing transactions during the operational year; (ii) was announced and became operational in different calendar years, so that announcement and opening effects can be separately identified for each data center; and (iii) became operational within 4 calendar years of announcement, so that the opening occurs within the 5-year event window and at least one subsequent event year is observed. Column 1 reports the overall post-announcement effect estimated using the main difference-in-differences specification (eqn. 2). Column 2 reports estimates from eqn. (SI 6). In this specification, the post-announcement coefficient captures the average differential change in housing prices after the announcement but before opening, and the post-opening coefficient captures the incremental change beginning in the opening year relative to the announcement-to-opening period. Standard errors are clustered at the census tract level and are reported in parentheses. Asterisks denote p-values: * <0.10, ** <0.05, *** <0.01.

Supporting Information References

[1] J Schroeder, et al., IPUMS national historical geographic information system: Version 20.0 [dataset] (2025).

[2] C Wing, SM Freedman, A Hollingsworth, Stacked difference-in-differences, (National Bureau of Economic Research), Technical report (2024).

[3] EA Stuart, Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* **25**, 1 (2010).

[4] D Hernandez-Cortes, K Meng, P Weber, The environmental costs and geography of US data center expansion. (2026).

[5] S Greenhill, Noise pollution and infant health, (University of California, Berkeley), Technical report (2026).