What street-level imagery can and cannot tell us about campus sustainability: a measurement study of 252 U.S. campuses

Cantay Caliskan, Hari Acharya, Nyaradzo Mararanje, Albert Mejooli, and Zhihan Zhang

University of Rochester, Rochester, NY, USA

October 30, 2026

Abstract

A university's sustainability commitments are encountered in classrooms and in the paths, landscapes, and travel options that shape campus life. We examine what imagery reveals about campus environments and how that evidence relates to university sustainability ratings.

We linked 1.84 million Google Street View panoramas and aerial imagery to 252 U.S. campuses, grouping photographs into two-month periods. We compared the environments visible from the street and from above, checked how repeated views, capture dates, and automated image interpretation affected the measurements, and tested whether imagery could reproduce STARS and QS Sustainability scores. We also propose a separate index of visible campus conditions.

Vegetation measures aligned across street-level and aerial images and varied with season. Errors in some automated measurements and incomplete repeat coverage constrained other comparisons. Averaged across models, imagery explained at most 10% of variation in STARS overall scores, including when entire states were excluded from training. Imagery-based QS models explained about 27%. Adding imagery to institutional information improved STARS predictions, with similar gains for an unrelated outcome. These findings show the limits of reading institutional sustainability from campus appearances and motivate a transparent Visible Campus Environment Index (VCEI) focused on the conditions imagery can credibly describe.

"It is highly meaningful for a university to go beyond education and research and take practical steps toward restoring nature."

—Ban Ki-moon.

Keywords: campus sustainability; sustainability ratings; street-level imagery; computer vision; machine learning; higher education

1 Introduction

In the early 1990s, Oberlin College's need for space became an opportunity to ask what a university building should teach. Students and architects explored ecological design in a yearlong course in

1992–93, followed by design charrettes open to the wider community. The resulting Adam Joseph Lewis Center combined solar technology, water purification, gardens, landscape restoration, and feedback on building performance. These features were intended to make environmental responsibility part of everyday education through the building's design and operation (Orr, 2021). The building itself became part of the lesson, connecting what students learned in class with how their institution used resources and cared for its surroundings. The project embodied Orr's (1993) argument in *Architecture as pedagogy*: buildings, landscapes, and daily routes teach through the resource choices and institutional values they express.

That lesson extends beyond campus grounds. Universities educate professionals, produce knowledge, and sustain communities with environmental consequences. Cortese (2003) connected teaching, re-search, operations, and community engagement within one institutional system, making sustain-ability a question of coherence across these activities. Commitments expressed in teaching gain practical weight when they also shape operations and the environments universities build. The same framework includes educational, organizational, and social achievements whose meaning ex-tends beyond the physical campus. Lozano et al.'s (2015) survey of 70 institutions found that commitments and implementation went together, yet efforts often remained compartmentalized. Together, these works motivate examining where commitments become observable conditions and how those conditions relate to broader institutional assessments.

Ratings organize such evidence at scale. AASHE's Sustainability Tracking, Assessment & Rating System (STARS) supports comparison and improvement among participating institutions (AASHE, 2024b); QS Sustainability and THE Impact Rankings provide global assessments (QS Quacquarelli Symonds, 2026; Times Higher Education, 2025). STARS draws on universities' own reporting, bringing together academic, engagement, governance, and operational achievements whose visibility ranges from physical infrastructure to educational practices. Critiques identify incomplete coverage, implicit norms, and limited comparability (Shriberg, 2002; Bullock & Wilder, 2016; Fischer et al., 2015; Findler et al., 2019; Boiocchi et al., 2023). STARS scores also track institutional type and size (Aldogan Eklund, 2026), while high ratings do not reliably accompany low greenhouse-gas emissions (Doocy et al., 2021). Campus observation could add evidence within a clearly bounded scope.

Imagery makes this possibility tangible. Adenle et al.'s (2020) review of 13 appraisal tools found limited spatial indicators and GIS integration. Google Street View (GSV) supports measurements of greenery, built form, perceived safety, and neighborhood demographics (Li et al., 2015; Seiferling et al., 2017; Naik et al., 2014; Gebru et al., 2017; Biljecki & Ito, 2021). Campus studies examine walking environments, restorative spaces, and active travel (Qin et al., 2025; Wu et al., 2025; Bai et al., 2022), although small samples and evaluation without campus grouping constrain some applications (Dileep et al., 2024). We ask: *how much information about sustainability ratings can be recovered from visible campus conditions, and what can an independent index of those conditions add?*

Three measurement problems intervene. Overlapping panoramas, repeated captures, and objects seen from multiple positions can inflate counts and overweight densely photographed streets (Krylov et al., 2018; W. Zhang et al., 2018; Fan et al., 2025). Capture dates also vary within campuses

and between neighboring panoramas (Curtis et al., 2013; Kim & Jang, 2023), while greenery carries seasonal bias (Zhao et al., 2025). Annual estimates and alignment with rating periods therefore re-quire scrutiny. Finally, detector performance may deteriorate on panoramas unlike curated training images, and equirectangular pixel shares require projection-aware interpretation.

We address these problems using 2.49 million panorama files, including 1.84 million linked to 252 U.S. campuses, NAIP aerial imagery, and sustainability scores. *Our first goal is to estimate existing scores from image-derived indicators*, evaluating predictions on held-out campuses and states and comparing imagery, institutional covariates, and their combination. Our indicators cover vegetation, buildings, roads, sidewalks, people, vehicles, and selected campus infrastructure. *Our second is to construct a transparent Visible Campus Environment Index (VCEI).* Its dimensions, directions, normalization, weights, and aggregation reflect what imagery can credibly measure. Construction remains separate from score estimation, with sensitivity analysis assessing index values and rankings.

We treat measurement decisions as results. Every measure is calculated from all panoramas (Version A) and with overlap and padding corrected (Version B), carrying both through descriptions, predictions, and index construction. Captures enter two-month windows without chaining; alternative windows test temporal sensitivity and annual feasibility. We ask how often the available imagery supports campus-year observations and comparisons with the dates represented by sustainability reports. Comparisons with overhead canopy, season, and screened detections assess validity. Actual images, extraction outputs, coverage maps, and timing summaries make the evidence and its gaps inspectable.

We searched widely for predictive accuracy, from ordinary and penalized regression through polynomial, spline, and kernel models to boosting, neural networks, and ensembles fixed in advance. Repeated-measurement models used every observation window. The models used imagery and institutional covariates, never another rating as a predictor. Expanding the search to 33 randomly tuned learners changed accuracy by at most 0.03; with the same learners, additional overhead features lowered predictive performance for every STARS outcome under state holdouts. Averaged across learners, imagery explained at most about 10% of STARS overall-score variation, with or without state holdouts. Operations and Transportation were estimated about as well as a curriculum-and-research score a camera cannot observe; Grounds was missed entirely. QS predictions were stronger (R^2 = 0.27 from imagery; 0.39 with institutional covariates), largely through institutional size and type. Signals came mainly from people, bicycles, vehicles, and photography patterns, not green-ery. Greenery can describe campuses with seasonal control; current custom detectors cannot reliably measure their targets, and recent coverage cannot support annual measurement or contemporaneous STARS matching.

The findings return us to Oberlin's lesson: campus environments matter, while institutional sustainability reaches beyond their appearance (Cortese, 2003). Even strong prediction would establish a statistical correspondence requiring interpretation and further evidence about its relationship to broader institutional sustainability. Building on Orr's attention to surroundings and Lozano et al.'s (2015) concern with implementation, we document what imagery can contribute and where its limits lie. Reporting this careful negative result helps prevent small predictive gains from being

mistaken for evidence that images capture institutional sustainability.

2 Literature review

Three strands of research bear on this study: how institutional sustainability is represented through indicators, what environmental evidence visual observation can supply, and how universities use assessment results. Read together, they show what image-based campus measurement can contribute and where its conclusions stop.

2.1 Sustainability indices, ratings, and assessment frameworks

University sustainability assessment has developed through instruments with different purposes and constructions. An index aggregates selected indicators, a rating classifies performance against stated criteria, and a ranking orders institutions within a comparison group. Each expresses choices about what sustainability means and how far it extends. In an early review of 11 cross-institutional tools, Shriberg (2002) argued that good tools should look beyond eco-efficiency to institutional processes and motivations.

Researchers have put these choices into practice in different ways. Lozano's (2006) Graphical Assessment of Sustainability in Universities (GASU) adapted sustainability reporting guidelines to universities and presents the results as a multidimensional profile that can be compared across institutions. Lukman et al. (2010) built a three-dimensional ranking index from research, educational, and environmental indicators, weighted them with the analytic hierarchy process, and applied it to 35 highly ranked universities. The two approaches mark the difference between displaying a profile of institutional activities and computing a single comparative measure.

Established programs also differ in scope and output. STARS is voluntary and self-reported, reviewed by AASHE staff, and valid for three years. It awards points across Academics, Engagement, Operations, and Planning & Administration, scores an institution as the percentage of applicable points earned plus bonus points, and publishes reports with recognition levels (AASHE, 2019, 2024b). Comparative reviews count it among the most comprehensive frameworks (Bullock & Wilder, 2016). Version 3.0 launched in June 2024 alongside version 2.2, so scores from the two structures coexist and are not strictly comparable (AASHE, 2024a). UI GreenMetric, by contrast, ranks universities worldwide; its 2026 edition, renamed the UI GreenMetric Sustainable University Rankings, has seven categories after adding Governance and Digitalization (Universitas Indonesia, 2026). Empirical comparisons therefore need to state the program, reporting period, and edition behind each score.

Comparative studies show why these distinctions matter. Lauder et al. (2015) trace GreenMetric's design choices to its intended users, goals, and practical constraints. Berzosa et al. (2017) applied four assessment tools to the same university and found that a strong result under one tool did not guarantee a strong result under the others. A single-institution comparison shows that tools can reach different diagnoses; it does not show how common such disagreement is across the sector.

Scope also shapes what a score means. Analyzing 1,134 indicators from 19 tools, Findler et al. (2019) found that effects beyond the institution received little attention and that internally available proxies were often used to assess impacts, and Alghamdi et al. (2017) likewise found that most frameworks emphasize environmental and operational themes. Only a few STARS credits concern things a camera can see: landscape management and biodiversity in version 2.2, ecologically managed grounds in version 3.0, commute modal split in both, and support for sustainable transportation in 2.2. For this study, then, visible campus conditions form a bounded environmental domain within a much broader institutional framework.

2.2 Visual evidence of environmental conditions

Visual measurement gives geographically explicit evidence about selected environmental features, which campus appraisal tools rarely use (Adenle et al., 2020). Street-level imagery offers one way in. The green view index (GVI) measures the share of vegetation in street-level views (Li et al., 2015), and deep segmentation models trained on ADE20K, Cityscapes, or Mapillary Vistas have replaced color thresholds (Zhou et al., 2017; Cordts et al., 2016; Neuhold et al., 2017; Cai et al., 2018). GVI describes what enters the camera's view, the eye-level greenery that overhead canopy maps miss. The two correlate weakly at single points (Larkin & Hystad, 2019) but more closely when many views are aggregated over an area (Stubbings et al., 2019), so reading one alongside the other requires attention to their different spatial meanings.

A related literature measures how environments are perceived. F. Zhang et al. (2018) trained a deep model on human ratings of street images to predict perceptions such as safety, liveliness, and beauty, and linked each perception to the objects in view; Naik et al. (2014) did the same for perceived safety. These outcomes are judgments about scenes. They support the study of environmental perception, but claims about actual safety or other substantive outcomes need separate evidence.

Campus studies show that these methods carry over to higher education. Qin et al. (2025) combined semantic segmentation, machine learning, and volunteer ratings to model perceptions of walking environments on four Chinese campuses from 15,596 Baidu street-view images. Their outcome is perceived campus quality rather than an institutional sustainability score, so whether visual campus indicators track broader assessments remains a separate empirical question.

Imagery has also been tied directly to university sustainability scoring. İncekara et al. (2022) used Sentinel-2B images of the main campuses of two Turkish universities to calculate scores for GreenMetric's Setting and Infrastructure category and compared them with the scores the universities had officially claimed. Fuentes et al. (2022) estimated the same category's indicators for a single campus from drone point clouds. These are direct precedents for deriving assessment information from imagery, but they concern one category and one or two campuses, not overall scores across a large sample of universities.

The reliability of visual indicators depends on how observations are made. Biljecki et al. (2023) found that measures of urban form and greenery vary with viewing direction, field of view, and image type, with greenery the least stable, although city-scale aggregates agree closely. How views

are sampled, whether as directional crops or whole panoramas, also changes GVI (Yin & Hirata, 2025; Chen et al., 2019). Coverage is uneven and socially patterned (Fry et al., 2020; Kim & Jang, 2023; Fan et al., 2025), and image quality and availability vary across places and years (Hou & Biljecki, 2022; Biljecki & Ito, 2021).

Counting objects adds a further problem: deciding when two detections show the same object. Multi-view methods triangulate rays from several panoramas, cluster their intersections, or learn to re-identify objects across views (Wegner et al., 2016; Krylov et al., 2018; W. Zhang et al., 2018; Laumer et al., 2020; Nassar et al., 2020). Without such a step, counts scale with the density of captures rather than with the density of objects.

Time is the other complication. Capture dates within a city spread over about seven years on average (Kim & Jang, 2023) and change within a few hundred meters along a street (Curtis et al., 2013). Historical captures make change analysis possible (Li, 2021), but seasonal differences in greenery are large: across 40 cities, season produced a mean absolute percentage error of 54% in GVI (Zhao et al., 2025; Han et al., 2023). Phenology varies with climate, so any calendar definition of leaf-on is an approximation (X. Zhang et al., 2003).

A recent review also limits how broadly novelty can be claimed. Yuan and Yu (2026) propose a green-campus assessment framework that pairs a comprehensive institutional structure with a spatial evidence module built on remote sensing and GIS, and they discuss comparability across campus types. They present it as a conceptual reference rather than a validated tool. Combining visual and institutional evidence is therefore an established research direction, and this study's contribution has to rest on its empirical tests and the validity of its measurements.

This work implies three levels of interpretation: the scene features that are observed, the environmental conditions inferred from them, and institutional sustainability more broadly. Moving from one level to the next requires both an argument and validation. A detected bike rack or waste bin is evidence that the facility exists, not that it is used or works well. Likewise, reading vegetation as an institutional achievement requires attention to climate, season, landscape management, and the scope of the measure. These distinctions guide the construction of the index in this paper.

2.3 How universities use sustainability assessments

Universities use assessments for internal diagnosis and comparison. Kamal and Asmuss (2013) compared four tools against the needs of the University of Saskatchewan and found STARS the most suitable for tracking sustainability across the whole institution, a choice that depended on what the university wanted to track. Comparing 12 tools, Fischer et al. (2015) showed that the choice of indicators sets implicit standards for what a sustainable university should address, with some fields dominating and others marginalized. Assessment frameworks therefore organize institutional attention as well as summarize performance.

Repeated assessment can support policy development and organizational learning. Lambrechts (2015) examined four AISHE assessments conducted between 2003 and 2010 in the business studies department of a Belgian higher education institution and described how they fed into departmental

and institution-wide policy. Later assessments also revealed stagnation and initiatives that were not sufficiently embedded in organizational structures. The case supports assessment as an input to policy, while showing that lasting implementation depends on continued organizational support.

Assessment can also inform educational and engagement programs. In interviews with administrators at 20 U.S. research universities, Lad and Akerlof (2022) found that sustainability literacy and culture assessments were undertaken mainly to earn STARS points but also to inform campus pro-grams. This evidence concerns one group of STARS credits and rests on administrators' accounts, so it offers no causal estimate of the effects of taking part in the rating.

University documents show strategic and public uses. The University of Louisville (n.d.) calls STARS its primary tool for measuring sustainability progress and ties improvement to its Pathway to Platinum plan, and Colorado State University (n.d.) lists its STARS Platinum rating among other sustainability distinctions on its admissions website. These examples document monitoring, target setting, and communication with prospective students; whether ratings affect applications or enrollment is a separate question.

Recent research examines how ratings and institutional performance relate. Combining external sustainability indicators for Italian universities with the UI GreenMetric ranking, Alberti et al. (2026) used Granger tests and dynamic panel models and reported a two-way relationship: rankings reflect performance, and institutions improve on what rankings measure while other dimensions receive less attention. These temporal associations do not by themselves identify the effect of a ranking. Studies of what accompanies STARS scores in the United States have likewise used institutional data, such as size, type, enrollment, endowment, and emissions, rather than imagery (Minutolo et al., 2021; Aldogan Eklund, 2026; Doocy et al., 2021).

2.4 Implications for this study

Existing research shows that imagery can describe campus environments, but provides less evidence about how those descriptions relate to institutional sustainability ratings. The closest applications estimate individual ranking categories for one or two campuses (İncekara et al., 2022; Fuentes et al., 2022), predict perceived campus quality (Qin et al., 2025), or develop conceptual frameworks (Yuan & Yu, 2026). Whether visible campus conditions explain differences in sustainability scores across universities remains an empirical question. Among the studies reviewed here, we identified none that examines this relationship across many universities while addressing view overlap, capture timing, and detector validity.[1]

Our two analyses address different parts of this question. Predicting existing scores tests how much information about institutional ratings can be recovered from campus imagery. Constructing a separate index asks which visible conditions the images can measure credibly, regardless of their association with those ratings. The index is therefore developed independently of the prediction models: agreement with an existing score is a result to investigate, rather than a criterion used to select its indicators or weights.

[1]The review draws on targeted web searches and is not an exhaustive review of bibliographic databases.

The measurement comparisons are central to both analyses. Versions A and B show how overlap correction changes campus indicators, predictions, and index values. Two-month observation windows and coverage assessments establish which places and periods can reasonably be compared. Checks against overhead canopy, season, and reviewed detections test whether the extracted features support their intended interpretation, while image examples and descriptive results allow readers to examine the underlying evidence. Annual comparisons require sufficient repeat coverage and alignment with score reference periods. These checks matter because a reliable measure of campus vegetation may still be a poor predictor of a university's broader sustainability rating.

3 Data

3.1 Campuses and outcomes

Campus polygons came from the Homeland Infrastructure Foundation-Level Data (HIFLD) Colleges and Universities Campuses layer (boundary vintage 2018), and we laid a 10 m grid of points inside each polygon. The analysis set has 252 campuses, listed in Table S15. Of these, 198 are the main campuses of institutions in the U.S. News 2026 top 100 national universities and liberal arts colleges; identifying them required correcting a fuzzy name match that had linked 374 unrelated or satellite campuses to ranked names. Another 41 campuses have a STARS report, and 13 more have a QS Sustainability score. Institutional covariates (total enrollment, Carnegie basic classification, and locale) came from IPEDS-derived institutional records.

STARS reports were taken from the AASHE public report page of each institution's most recent report (submitted 2022–2025). For each analysis campus with a numeric score (n = 130; 92 under STARS 2.2 and 38 under STARS 3.0), we collected the subcategory points shown on the dated report page. Sub-scores are public for 109 reports. The other 21, all STARS 2.2-era reports submitted between November 2022 and September 2023 whose ratings have since expired, show no subcategory points. We checked every collected report by recomputing the overall score from its subcategory points, and the result matched the page to within 0.005 points in all 109 cases. Five page scores differed slightly from an earlier download of the scores (largest difference 0.48 points); we use the page values. The second outcome is the Sustainability lens score of the QS World University Rankings 2025, available for 113 analysis campuses (methodology: QS Quacquarelli Symonds, 2026), which we call QS Sustainability.

3.2 Street-level imagery

The corpus holds 2,494,281 equirectangular GSV panoramas stitched from Street View tiles. Each comes with a panorama identifier, location, vehicle heading, and capture year and month; no capture day is available. The project's pipelines had already processed every panorama. Vegetation came from Mask2Former (Swin-L; ADE20K tree, grass, and plant classes) with an HSV filter, both as a plain pixel share and as a share weighted by depth from DPT-Large (Cheng et al., 2022; Zhou et al., 2017; Ranftl et al., 2021). Road, sidewalk, and building shares came from DeepLabV3+ trained

on Cityscapes (Chen et al., 2018; Cordts et al., 2016), and persons and vehicles from a COCO-trained YOLO model (Lin et al., 2014). Bike racks (YOLOv10-s), garbage bins and bus stops (YOLOv8-s), and EV chargers (YOLO11-s segmentation) were detected with models trained on 522–1,020 Roboflow images per class and run with sliced inference on 2,048-px tiles at a confidence threshold of 0.80 (Wang et al., 2024; Jocher et al., 2023; Jocher & Qiu, 2024; Akyon et al., 2022). Supplementary Note S2 lists every image-processing model with its settings (Table S9) and reports how the four detectors were trained and how they performed on their own test sets (Tables S10 and S11). DeepLabV3+ was run on the full canvas, including the black padding of padded files, which it labels mostly as road and building; Section 4.3 describes the correction. Person and vehicle counts (car, truck, bus, and motorcycle) use a confidence threshold of 0.50.

3.3 Overhead imagery and climate

NAIP natural-color aerial imagery from 2018–2023, flown mainly in the growing season (U.S. Geo-logical Survey, 2026), was classified into eight land-cover classes by a Mask2Former model trained on OpenEarthMap (Xia et al., 2023). From these we use the shares of tree canopy, all vegetation, roads, buildings, other paved surfaces, and total impervious cover in each campus polygon. For each campus we use the May–September acquisition closest in time to its street-level window, or the closest acquisition from any month for the 16 campuses without one; one campus has no NAIP acquisition. The NAIP imagery in this collection has no near-infrared band, so NDVI could not be computed.

For the wide search (Section 4.5) we also built the full overhead block that the segmentation supports, averaged over all of a campus's acquisitions (2018–2023; median three). It contains the shares of all eight classes and of vegetation and impervious cover; three ratios (vegetation to impervious, tree to impervious, and building to vegetation) and the road share of impervious cover; the Shannon diversity and evenness of land cover; the largest-patch share of six classes and the patch density of eight; the campus area; and, as image metadata, mean classification confidence and entropy, the share of low-confidence pixels, the mean brightness of each band, and the number of acquisitions, for 38 features in all (Table S8). Because NAIP is flown state by state, the image metadata differ by state. State explained 32–47% of the variance in mean band brightness across campuses and 41% of the variance in the number of acquisitions, against about 15% expected by chance among 39 states.

Monthly temperature and precipitation from NOAA nClimGrid (Vose et al., 2014a, 2014b) and the 1991–2020 normals (Palecki et al., 2021) could describe capture-month conditions and campus climate, but we did not extract them for this study, so season is represented by the capture month.

4 Methods

Figure 1 summarizes the pipeline.

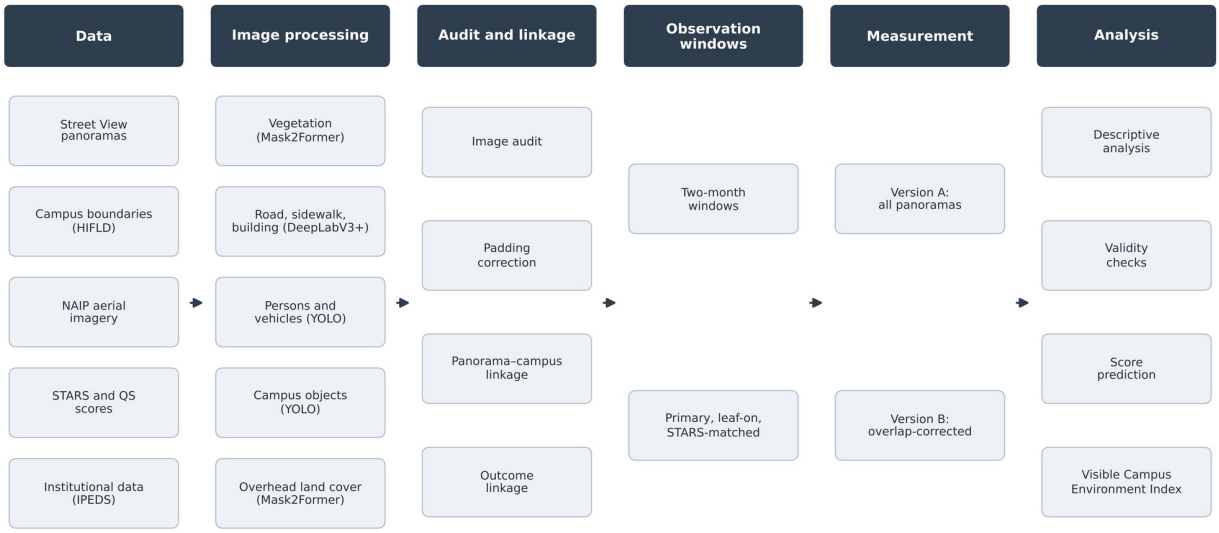

Figure 1: Analysis pipeline.

4.1 Image audit and linkage

We opened the panorama files and classified each canvas (Table 1). Three artifacts of the tile download matter. First, two-thirds of the 4,096-px canvases (1,232,541 files, half of the corpus) hold a 3,328 × 1,664 px panorama padded with black, in which columns 3,328–3,583 repeat columns 0–255, so a 27.7° sector appears twice. Second, most valid 16,384-px canvases (61%) are also padded, though without a repeated strip. Third, 74,416 files (3.0%) are entirely black. These are failed downloads of capture dates for which the requested zoom level does not exist, and we excluded them from both versions. Checking sample images confirmed that the center column of the true panorama faces the stored vehicle heading. That fixes the conversion from a horizontal pixel position to a compass bearing: bearing = heading + (x / W − 0.5) × 360°, where W is the width of the true panorama.

A panorama was linked to a campus if it lay within 20 m of the campus's 10 m grid. We measure coverage against coverable points, meaning grid points within 20 m of a panorama captured at any date, because most of a campus (roofs, lawns, and building interiors) can never be photographed from a street car.

4.2 Observation windows

Because capture dates are known only to the month, we define observation periods in calendar months. Our primary rule forms windows of at most two consecutive calendar months, chosen greedily: the two-month span that covers the most 20 m units is taken first, its months are removed, and the process repeats. Each window is a fixed span anchored at its first month, so a run of consecutive months can never chain into a longer period, and no month belongs to two windows (see Section 5.2). We compare this rule with calendar months, fixed two-month blocks, calendar quarters,

and meteorological seasons. Each window records its months, a reference month (the unit-weighted median), and its coverage. For each campus we then define three windows: a primary window (highest coverage), a leaf-on window (the best-covered window entirely within May–September), and a STARS-matched window (among windows covering at least 20% of coverable points, the one closest to the month of the STARS submission).

4.3 Version A and Version B

Version A uses every valid panorama in a window with equal weight. Campus values are means of panorama shares, and detections are counted per panorama.

Version B makes four corrections. (i) For area shares, the campus is divided into 20 m sampling units and only the panorama nearest each unit's center is kept (ties go to the later capture), so densely driven streets do not dominate. (ii) Detections centered in the repeated 27.7° strip of padded 4,096-px canvases are removed. (iii) The remaining detections of each class are converted to bearing rays. Rays from panoramas within 30 m of each other are intersected when the intersection lies 1.5–30 m from both cameras at an angle of at least 8°, and the intersections are clustered (DBSCAN, 3 m) into objects, which are expressed per 100 sampling units (Figure 2). (iv) Road, sidewalk, and building shares were computed by DeepLabV3+ on the full canvas, and an entirely black file is scored 41% road and 56% building, so padding inflates these shares. Comparing padded and full canvases captured at the same 20 m units on different dates (108,213 units at 4,096 px and 55,509 at 16,384 px), we estimate that the model labels 70% of padding pixels as road (63% at 16,384 px), 5% as building (24%), and almost none as sidewalk. Version B removes this contribution from each padded panorama as share = (observed − (1 − c) k) / c, where c is the true-image share of the canvas and k is the class rate. After the correction, padded and full canvases have nearly the same mean road (0.41 and 0.42) and the same building (0.07) shares. Version A keeps the stored shares. Both versions use the same windows, campuses, models, and cross-validation folds.

4.4 Measurement validity checks

We compared street-level greenery with NAIP tree canopy across campuses, and within campuses between a leaf-off primary window and the campus's best leaf-on window. For the special objects (bike racks, bus stops, EV chargers, and garbage bins) we first tabulated detection rates and box sizes by image resolution over all detections. We then screened 96 detections at confidence $\geqslant$ 0.80 in panoramas linked to analysis campuses, 24 per class: 16 drawn as two per class from each resolution (shown in Section 5.4) and 80 drawn at random from each class's detections, so that 70 of the 96 came from 16,384-px files, roughly as in the population. Each crop was judged as showing the named object or not. Because each detection was judged once rather than by independent labelers, we report it as a screening, with Wilson intervals (Wilson, 1927). The detectors' training data and test-set accuracy are reported in Supplementary Note S2 (Tables S10 and S11).

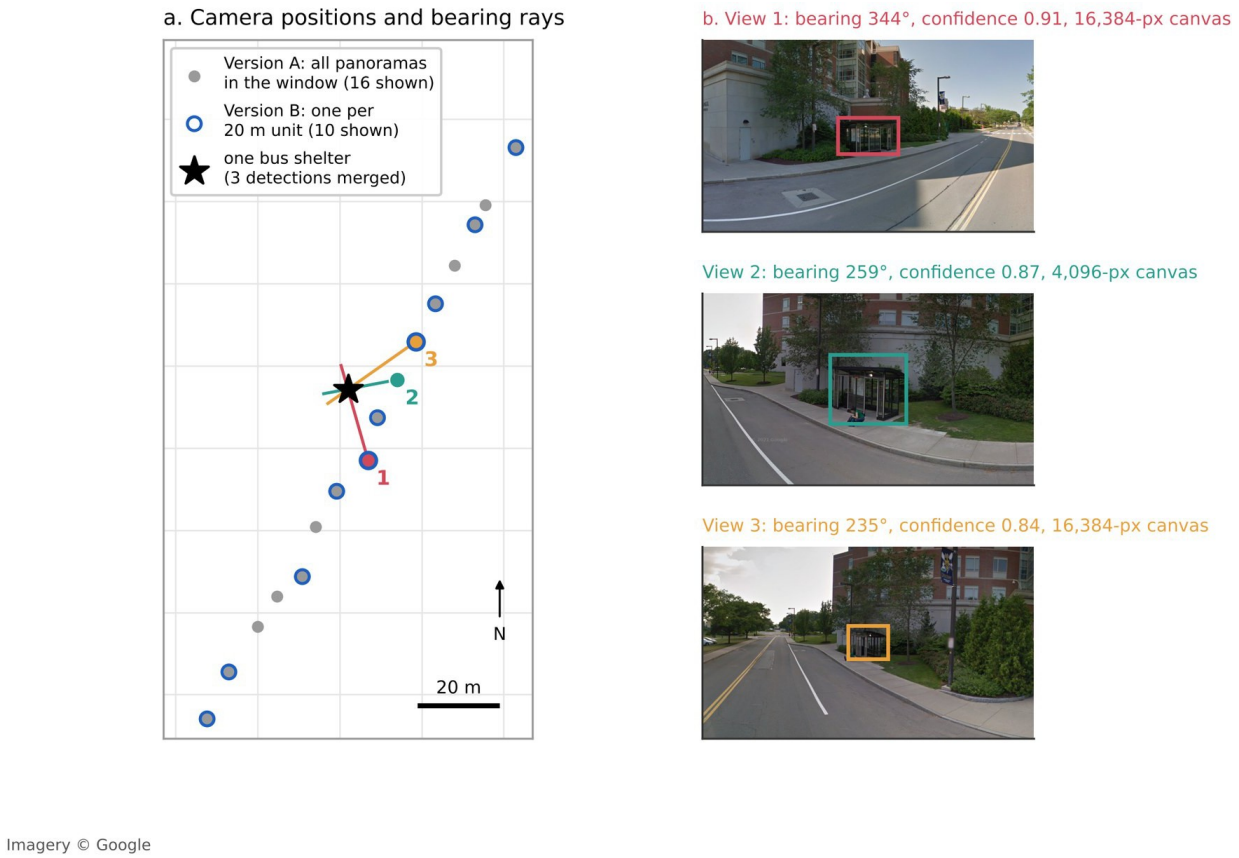

Figure 2: Overlap correction at the University of Rochester, July 2015. (a) Version A keeps every panorama in the window and Version B one per 20 m unit; three sightings of one bus shelter, from panoramas 12–20 m away, merge where their bearing rays intersect. (b) The three detections; the match was checked visually, not by an independent labeler.

4.5 Estimating sustainability scores

We model five STARS outcomes and one QS outcome. The STARS outcomes are the overall score; three sub-scores that imagery could plausibly reflect, namely Operations (all operational subcategories), Transportation, and Grounds (the 2.2 Grounds subcategory or the 3.0 Ecologically Managed Grounds credit); and the Curriculum and Research subcategories, which serve as a negative control because imagery has little reason to reflect them. Sub-scores are percentages of available points. The sixth outcome is the QS Sustainability score. No predictor is derived from any sustainability rating. Models of a STARS score never use QS or other STARS scores, models of QS never use STARS, and the STARS version (2.2 or 3.0) is not used either, even though the two versions score on somewhat different scales. The baseline is the training-fold mean.

For each campus and window, every street-level measure is summarized over the window's Version B panoramas. The summaries are the mean of each share (green view, tree, grass, and depth-weighted green view; padding-corrected road, sidewalk, and building shares; and the sidewalk-to-road ratio), the mean of each count (vehicles, persons, bicycles, cars, trucks, and buses per panorama), and, for the four special objects, detections per panorama together with merged objects and presence per 100 sampling units. With the NAIP shares and the window's imagery-quality covariates (coverage, number of units, share of high-resolution files, reference year, and leaf-on flag), they form a compact set of 41 features. A full set of 98 features adds the within-window distributions: the SD and the 10th, 50th, and 90th percentiles of each share, and the share of panoramas with at least one detection of each count.

Imagery comes from three sources: the primary window of Section 4.2; the window closest to the STARS submission (STARS outcomes only); and the average of window-level measurements over every window that covers at least 20% of a campus's coverable points. In this multi-window design, windows are measured separately and averaged as repeated measurements rather than chained. Feature sets are institutional covariates (log enrollment, Carnegie class, locale, and census region), imagery, and the two combined. To compare the two measurement versions, two further sets hold only the ten measures of the study design (green view; road, sidewalk, and building shares; vehicles and persons per panorama; and bike racks, garbage bins, bus stops, and EV chargers) in the primary window, computed either from all panoramas with stored shares and raw detections (Version A) or from Version B. The STARS-matched window uses the date of a campus's STARS submission to decide which images to use. Its capture year would therefore carry the submission date, and with it the STARS version, so the year is left out of those models; no rating enters any model.

To find the most accurate estimates the data allow, we evaluated 27 learners from eight families on every outcome, window, and feature set (Table S12 lists each learner with its settings). The linear models were ordinary least squares on a dozen window means, ridge, lasso, elastic net, Bayesian ridge, Huber regression, partial least squares, principal-component regression, and ridge after univariate screening. The polynomial models expanded screened features to degree 2 or 3 with ridge or lasso penalties, and the additive model fitted cubic B-splines to each feature with a ridge penalty, a form of generalized additive model. The kernel methods were kernel ridge with radial-basis and polynomial kernels, support-vector regression with radial and linear kernels, and a Gaussian pro-

cess on ten principal components. The remaining families were k-nearest neighbors, tree ensembles (random forest and extremely randomized trees), gradient boosting (scikit-learn, histogram-based, XGBoost, LightGBM, and CatBoost), and a neural network with one hidden layer (Pedregosa et al., 2011; Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018; Hastie et al., 2009). Numerical features were median-imputed and Yeo–Johnson transformed inside each training fold. Each learner was tuned by 5-fold cross-validation within each outer training set (boosting and forests used fixed small-data settings) and then predicted the outer test fold. Because the best of 27 cross-validated R^2 values overstates the accuracy of any procedure that picks it (Cawley & Talbot, 2010), we also report four pipelines fixed in advance, whose outer-fold accuracy is honest: the learner with the best inner-CV R^2 in each outer training set (nested selection), the average of the five learners with the best inner-CV R^2, stacking with non-negative weights fitted to inner out-of-fold predictions (Wolpert, 1992; Breiman, 1996), and the average of all learners.

A campus typically has eight two-month windows that each cover at least 20% of its coverable points and at least 20 units, and each window is a noisy look at the same campus. In the window-augmented models, every such window is a training row carrying its campus's score. Rows are weighted so that each campus carries the same total weight, and a campus's prediction is the unit-weighted mean of the predictions for its windows, in the spirit of multiple-instance learning (Dietterich et al., 1997). Outer and inner folds are drawn over campuses, so all windows of a test campus are held out together. We evaluated nine learners in this design (ridge, elastic net, kernel ridge, support-vector regression, k-nearest neighbors, random forest, extremely randomized trees, LightGBM, and CatBoost; Table S13) along with the four pipelines.

We then checked whether a wider search, or more of the overhead imagery, would find accuracy that this search had missed. This wide search used 33 learners (31 in the window-augmented design), each tuned by random search over wide ranges (Bergstra & Bengio, 2012; Table S14). It added automatic relevance determination (Tipping, 2001), lasso fitted by least-angle regression with an information criterion (Efron et al., 2004), orthogonal matching pursuit, median regression (Koenker & Bassett, 1978), kernel ridge with a Laplacian kernel, *ν*-support-vector regression (Schölkopf et al., 2000), and AdaBoost (Freund & Schapire, 1997), and its Gaussian process chose among radial-basis, Matérn, rational-quadratic, and linear-plus-radial-basis kernels (Rasmussen & Williams, 2006). Median regression and the Gaussian process were used with one row per campus only, and forests and boosting were tuned rather than fixed. For each learner we drew 20 settings at random (8–15 for slow learners; 12 and 5–6 in the window-augmented design) over penalties, kernel widths and degrees, numbers of components or selected features, numbers of neighbors, tree depth, leaf size and feature subsampling, learning rate and number of boosting rounds, and network size and weight decay. Each setting was scored by campus-level 5-fold cross-validation inside the outer training set, and the best was refitted on the whole training set. A run evaluated about 5,500 settings (3,000 in the window-augmented design). Where the features outnumbered the campuses of an inner training set, the least-angle lasso could not be fitted and was left out of the pipelines. The feature sets were those of the multi-window model (and, for the STARS overall score, of the window-augmented model); the same with the full overhead block (Section 3.3) in place of the six NAIP shares, in both designs; the full overhead block without its image metadata (one row per campus); and the full

overhead block with institutional covariates (one row per campus, and window-augmented for QS Sustainability). The pipelines were those above plus the median of all learners. Because of its cost, the wide search was evaluated only on the first random 5-fold split of the earlier search and on the state-grouped split. Cross-validated R^2 varies between random splits by as much as the differences at stake (0.15–0.27 across the three splits for QS Sustainability from multi-window imagery), so we compare the two searches on these identical splits.

Outer evaluation used 5-fold cross-validation repeated three times, plus 5-fold cross-validation grouped by state as a spatial check (Roberts et al., 2017). Every learner and feature set saw identical folds, and all preprocessing, tuning, selection, and ensemble weights were fitted inside training folds (Varma & Simon, 2006; Kapoor & Narayanan, 2023). Out-of-sample R^2 is computed against the training-fold mean and averaged over repeats, and what imagery adds to institutional covariates is the difference between the two feature sets on the same folds. For the window-augmented models, we built a null distribution by reassigning the imagery of whole campuses (all their windows and their NAIP shares) to other campuses at random, keeping scores and institutional covariates in place, and refitting three learners (ridge, kernel ridge, and extremely randomized trees) and their average on the first random split (50 permutations). With institutional covariates in the model, the same reassignment tests what imagery adds to them (20 permutations). The contribution of each block of features (greenery; road, sidewalk, and building; vehicles, persons, and bicycles; special objects; NAIP land cover; and capture metadata) is the drop in out-of-fold R^2 when that block is permuted jointly across the held-out campuses of each fold, with all windows of a campus moving together (mean of 10 permutations per fold). To show how well the measures fit in the classical sense, we also report in-sample ordinary least squares fits (R^2, adjusted R^2, overall and partial F-tests, and standardized coefficients with HC3 standard errors) of the ten measures, of all imagery features, and of institutional covariates. With about 100–130 campuses, cross-validated R^2 values are themselves uncertain (Varoquaux, 2018), and the search involves many comparisons, so we report the spread across repeats and give most weight to the pipelines fixed in advance.

4.6 Visible Campus Environment Index (VCEI)

The VCEI measures environmental conditions visible from the campus street network and from above in a stated window. It is not a measure of institutional sustainability, and it was not tuned to STARS or QS. It has three dimensions: greenery (street green view, street tree share, NAIP tree canopy, and NAIP vegetation), pedestrian space (sidewalk share and sidewalk-to-road ratio), and car dominance (vehicles per panorama, street road share, and NAIP roads and other paved surfaces, all reversed). Indicators are normalized with fixed goalposts, the 5th and 95th percentiles of the Version B primary-window distribution (Table S4), and clipped to [0, 1], so a campus's value does not change when other campuses enter or leave the sample (OECD and European Commission Joint Research Centre, 2008). Indicators are averaged within dimensions with equal weights, and the dimensions are combined by a geometric mean (floor 0.01), which limits how far strong greenery can compensate for heavy car dominance (Munda & Nardo, 2009); the arithmetic mean is reported for comparison. We assessed uncertainty with 500 Monte Carlo draws that vary the dimension

weights (Dirichlet, $\alpha = 1$), drop one indicator at random in about 45% of the draws, and alternate the aggregation rule, which yields 90% rank intervals (Saisana et al., 2005; Paruolo et al., 2013; Becker et al., 2017). Street-level road and sidewalk shares enter in their padding-corrected Version B form.

5 Results

5.1 What the corpus contains

Of 2,494,281 panorama files, 2,419,587 were valid images, and 1,835,308 were linked to the 252 analysis campuses (Table 1). Half of all files are padded 4,096-px canvases carrying the duplicated 27.7° sector (Figure S2). Resolution followed the download phase rather than the scene: 24.7% of files are 16,384 px wide, including the 74,416 black failures, and 75.3% are 4,096 px wide. Padding follows capture year. Nearly all panoramas captured before 2017 are padded, and the switch to full canvases happened over 2017–2019 (Figure S3). Resolution mostly does not: about a quarter of each year's files are high-resolution, as are a median 24% (IQR 22–25%) of the files in each campus's primary window. The exception is 2007–2009, when most high-resolution requests failed and returned blank files, so the 35 campuses whose primary windows fall in 2007–2008 are measured almost entirely from 4,096-px images.

Table 1: Panorama corpus and linkage.

Canvas type or step	Files	Share	Content and handling
4,096 px, padded	1,232,541	49.4%	3,328 × 1,664 true image, repeated 256-px strip, black padding; strip detections removed and road, sidewalk and building shares corrected for padding in Version B
4,096 px, full	646,067	25.9%	complete 360° image
16,384 px, padded	331,496	13.3%	13,312 × 6,656 true image, black padding; road, sidewalk and building shares corrected for padding in Version B
16,384 px, full	209,483	8.4%	complete 360° image
16,384 px, entirely black	74,416	3.0%	failed download; excluded
Unreadable	278	< 0.1%	excluded
Valid panoramas	2,419,587		
… with location, heading and capture month	2,419,563		
… within 20 m of any downloaded campus grid	2,275,967		910 campuses
… linked to the 252 analysis campuses	1,835,308		198 ranking targets + 41 other STARS + 13 other QS campuses

5.2 Timing and observation windows

The two-month anchored rule (Figure 3) produced a median of 18 windows per campus (IQR 11–27). The best window covered a median 78% of coverable points (IQR 71–87%) in Version A and 75% in Version B, compared with 75% for calendar months, 76% for fixed two-month blocks, and 78% for calendar quarters (Table S1). Moving from one month to two gains about four percentage points of coverage, while a third month adds little more and lengthens the period a window spans. For 12.3% of campuses, the best window straddled the May or September boundary of the calendar leaf-on period.

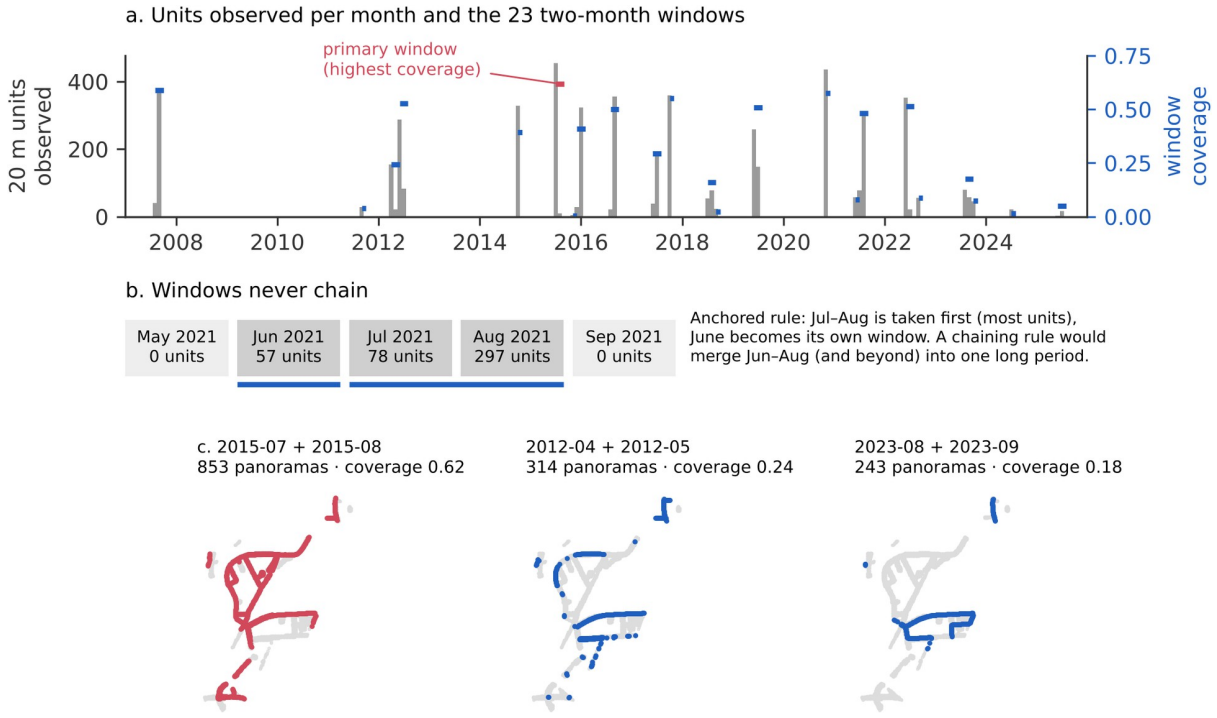

Figure 3: Observation windows at the University of Rochester. (a) 20 m units observed per month and the 23 two-month windows with their coverage. (b) Windows are fixed spans of at most two calendar months and never chain. (c) Panoramas in three windows over the campus's coverable points.

Annual campus measurement is not supported for most campus-years (Figure 4a). Of 4,788 possible campus-years (252 campuses × 2007–2025), 1,024 (21.4%) had a window covering at least half of the coverable points, and 1,853 (38.7%) had one covering a quarter. Coverage peaked in 2017–2019, when 41–48% of campuses had a window above 50%, and then fell to 11.1% in 2023, 2.0% in 2024, and 0.4% in 2025. As a result, the best-covered imagery precedes the STARS reports by a median
9.4 years (IQR 6.4–13.0; 141 campuses with a STARS report, including 11 reporters without a score; Figure 4b). Among windows covering at least 20% of points, the one closest to each STARS submission lies a median 1.8 years earlier (IQR 0.8–3.3) and covers a median 38% of points.

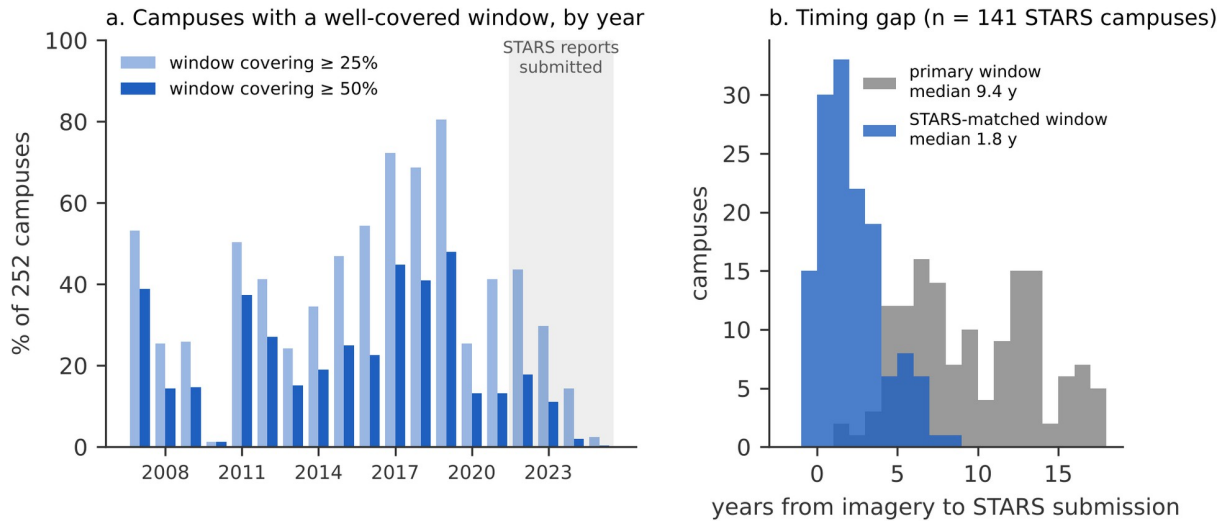

Figure 4: Annual coverage and timing. (a) Share of the 252 campuses with a two-month window covering at least 25% or 50% of coverable points, by year. (b) Years between the imagery window and the STARS submission, for the primary and STARS-matched windows.

5.3 What the measures look like

The 252 campuses are mostly research universities and liberal arts colleges in urban settings (Table 2), spread over 38 states and the District of Columbia, with concentrations in the Northeast and Midwest (Figure 5). Primary windows have a median reference year of 2014 (IQR 2011–2018), and 66% fall in May–September. A primary window contains a median of 522 panoramas in Version A and 302 sampling units in Version B.

Table 2: Campus sample and imagery per campus.

Section	Characteristic	Value
Campuses	Analysis campuses	252
	Ranking targets / other STARS-scored / other QS-scored campuses	198 / 41 / 13
Carnegie classification	Doctoral, very high research (R1)	96 (38%)
	Doctoral, high research (R2)	33 (13%)
	Doctoral/professional	7 (3%)
	Baccalaureate, arts & sciences	103 (41%)
	Other or missing	13 (5%)
Census region	Northeast	86 (34%)
	Midwest	52 (21%)
	South	66 (26%)
	West	48 (19%)
Locale	Urban	140 (56%)
	Suburban	60 (24%)
	Town	32 (13%)
	Rural	7 (3%)
	Missing	13
Size	Total enrollment, median [IQR]	7,643 [2,003–26,292]
Outcomes	STARS report with numeric score	130
	of which STARS 2.2 / 3.0	92 / 38

Section	Characteristic	Value
	with public sub-scores (Operations, Transportation, Grounds)	109
	report submission years	2022–2025
	QS Sustainability score	113
Imagery	Campus grid points (10 m), median [IQR]	8,138 [4,862–18,564]
	Share of grid points within 20 m of any panorama, median [IQR]	0.28 [0.19–0.37]
	Valid linked panoramas per campus, all dates, median [IQR]	3,430 [1,335–9,494]
	20 m sampling units ever observed, median [IQR]	433 [233–1,065]
	Two-month observation windows per campus, median [IQR]	18 [11–27]
Primary window	Coverage of coverable grid points, Version A, median [IQR]	0.78 [0.71–0.87]
	Coverage, Version B, median [IQR]	0.75 [0.68–0.84]
	Panoramas, Version A, median [IQR]	522 [271–1,179]
	Panoramas, Version B (one per unit), median [IQR]	302 [168–666]
	Reference year, median [IQR]	2014 [2011–2018]
	Share of windows in leaf-on months (May–Sep)	66%
	Share of panoramas from 16,384-px canvases, median [IQR]	0.24 [0.22–0.25]
	Usable (coverage ⩾ 0.10 and ⩾ 20 units)	249

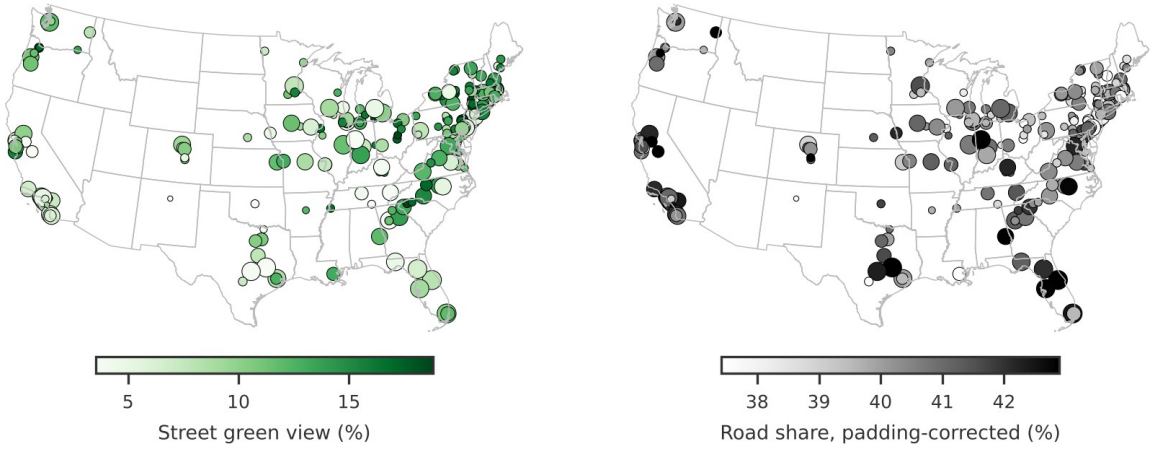

Figure 5: The analysis campuses. Color shows street green view (left) and padding-corrected road share (right) in each campus's primary window (Version B); marker area is proportional to enrollment. 249 campuses with a usable primary window.

Figures 6–8 show what the extracted measures look like in the images. Figure 6 reproduces six of 13 panoramas drawn at random (all 13 are in Figure S1). Figure 7 shows, for each of five measures, a randomly drawn panorama near the 5th, 50th, and 95th percentiles of all Version B panoramas. High green view turns out to be a tree-shaded parking area, high building share a street between a

residence hall and a parking garage, and the 95th percentile of vehicles a parking lot. The panorama at the 5th percentile of green view, scored at 1%, shows a large tree, a reminder of the segmentation's limits. Figure 8 shows person and vehicle detections on four panoramas from four strata.



Figure 6: Six randomly drawn panoramas. Full 360° views with padding removed and the vegetation shares used in the analysis; all 13 are in Figure S1.

Within campuses, the measures vary along streets in ways that campus means hide (Figure 9). At the University of Chicago, green view is high along a few planted streets and low elsewhere, while vehicle counts peak on the through streets. At Stanford, the campus core is greener than the open roads that cross the outer campus. At Colgate, a single June 2011 window observed only 97 units, so the campus value rests on a few roads. Coverage of this kind, and not only a campus's character, shapes each campus value.

Across campuses, the measures differ greatly in how much they vary (Table 3; Figure 10). Street green view ranged from 3.5% to 18.8% between the 5th and 95th percentiles (mean 10.7%, about seven-tenths of it trees), and NAIP canopy from 7% to 47%. Once corrected for padding, road share

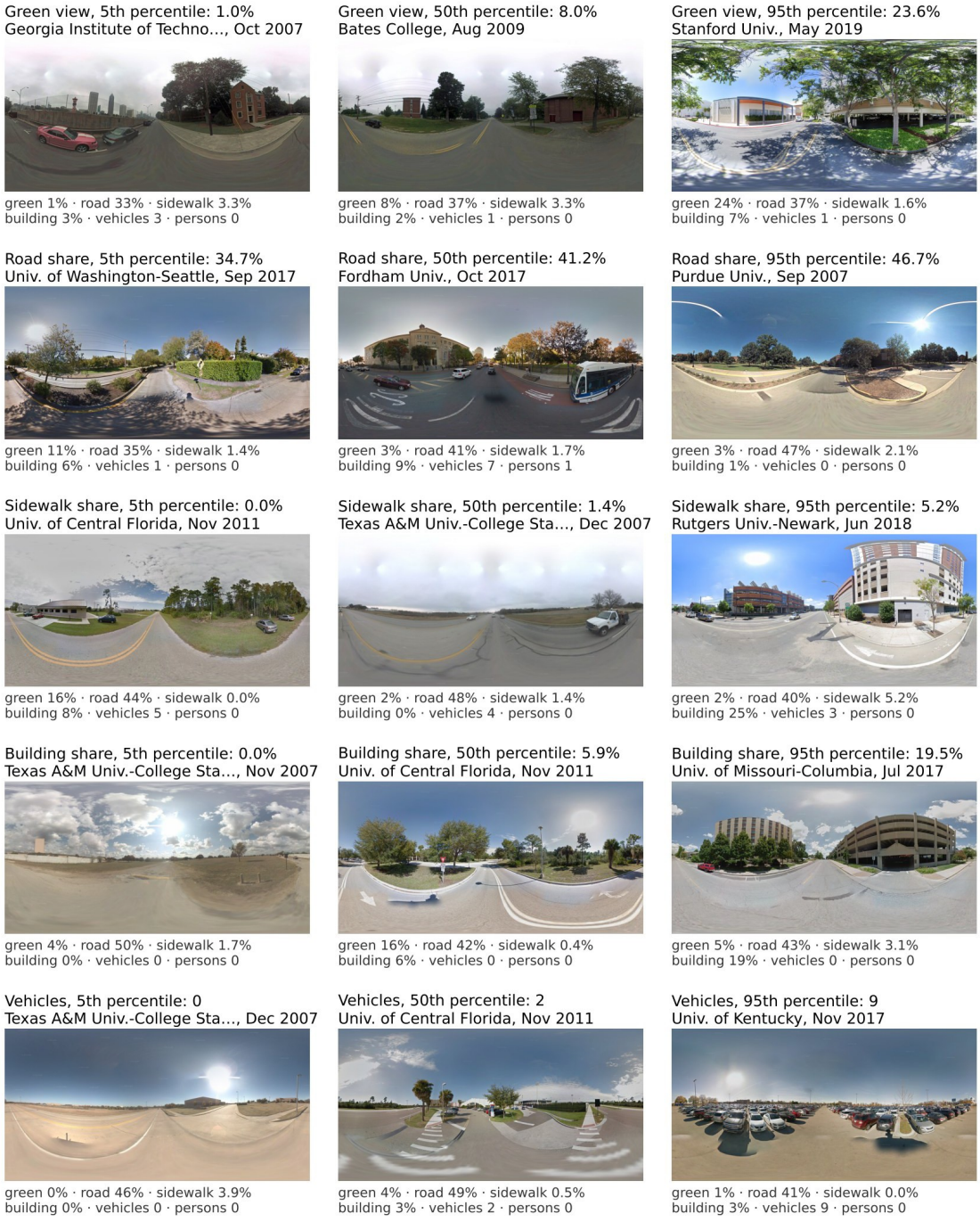

Figure 7: Panoramas at the 5th, 50th, and 95th percentiles of five measures. Each was drawn at random from the Version B primary-window panoramas within one percentile of the target (4,096-px canvases); every measure is listed below each image.

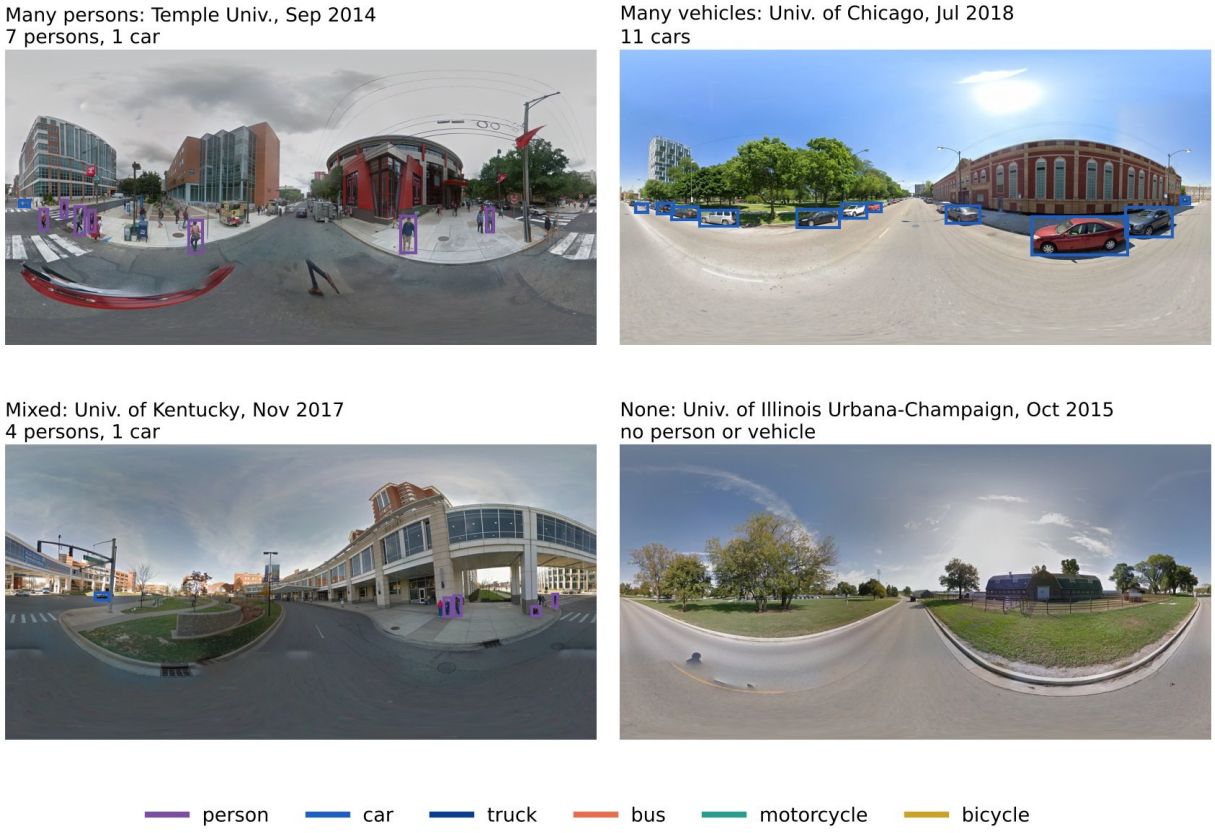

Figure 8: Person and vehicle detections. YOLO with COCO classes at confidence $\geq$ 0.50, on four randomly drawn panoramas from four strata: at least four persons; at least eight vehicles; one to three vehicles and at least one person; and no detection at confidence 0.25.

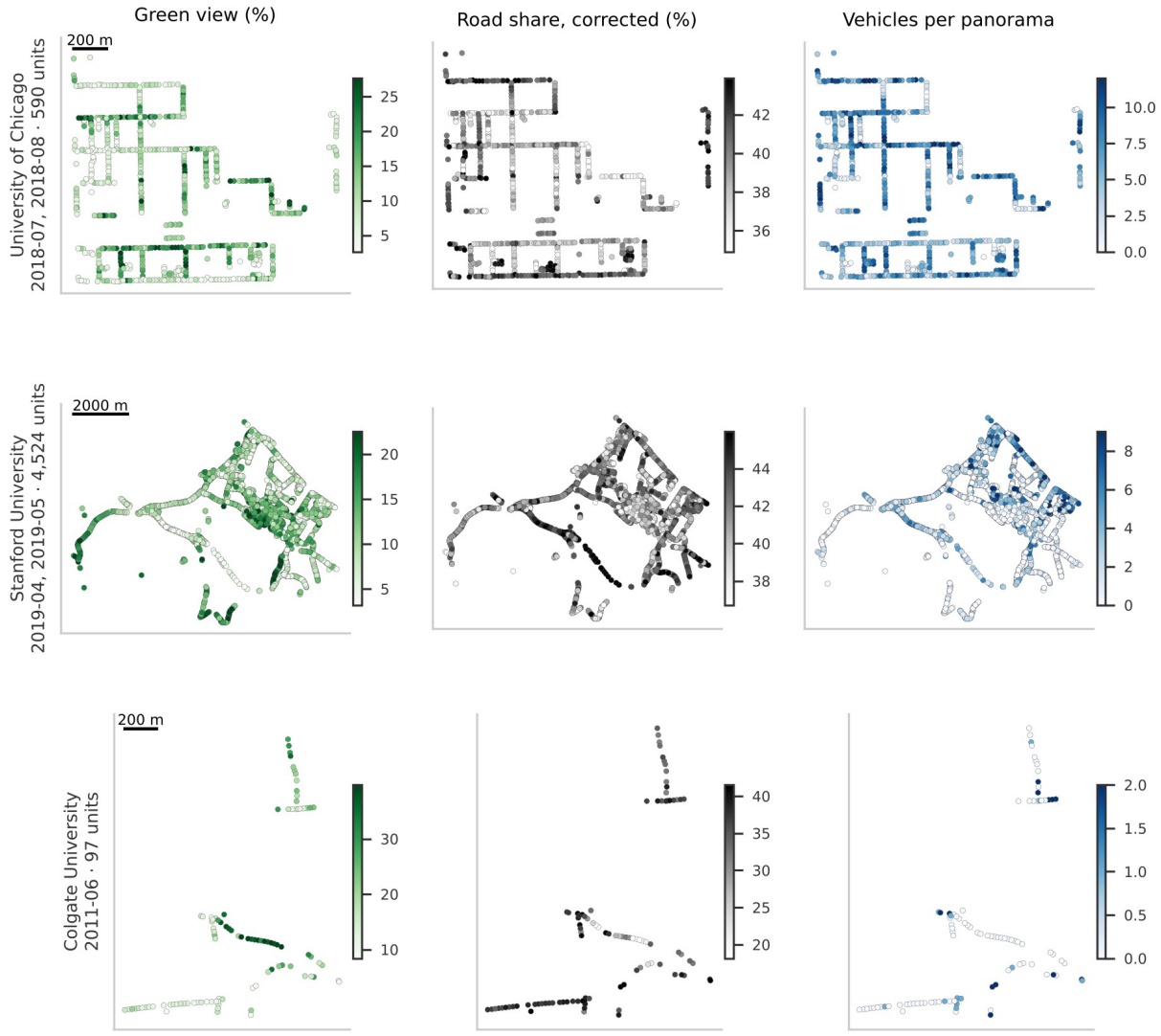

Figure 9: Within-campus variation at three campuses. Version B panoramas in the primary window, colored by green view, padding-corrected road share, and vehicles per panorama (color scales clipped at each campus's 5th and 95th percentiles).

varied little (37.4–42.9%; coefficient of variation 0.04), because every panorama is taken from a road, and sidewalk share was small (0.9–3.1%). Vehicles per panorama ranged from 0.6 to 5.6. Persons were rare and highly skewed (median 0.06 per panorama; coefficient of variation 1.6), and detected bicycles were absent from 57% of campuses. In Version A the stored road share was 6.8 points higher on average and the sidewalk share lower, which is the padding bias described in Section 4.3.

Table 3: Distribution of campus-level measures.

Measure	Mean (SD)	5th	25th	Median	75th	95th	Campuses with zero	Version A mean
Street green view (%)	10.7 (4.8)	3.5	7.0	10.7	13.9	18.8	0%	10.7
Street tree share (%)	7.7 (4.0)	1.8	4.9	7.3	10.4	14.3	0%	7.7
Street grass share (%)	2.6 (1.7)	0.3	1.1	2.5	3.8	5.7	0%	2.6
Green view, depth-weighted (%)	7.2 (3.4)	2.3	4.6	6.9	9.5	12.7	0%	7.2
Road share (%)	40.2 (1.7)	37.4	39.4	40.3	41.2	42.9	0%	47.0
Sidewalk share (%)	1.9 (0.7)	0.9	1.5	1.9	2.4	3.1	0%	1.5
Building share (%)	7.0 (4.2)	1.0	4.8	6.8	8.4	15.1	0%	7.4
Sidewalk / (sidewalk + road)	0.05 (0.02)	0.02	0.04	0.05	0.06	0.07	0%	0.03
Vehicles per panorama	2.7 (1.5)	0.6	1.6	2.5	3.4	5.6	0%	2.7
Persons per panorama	0.13 (0.21)	0.00	0.03	0.06	0.14	0.44	6%	0.13
Panoramas with a person (%)	8.0 (9.2)	0.0	2.3	4.7	9.9	26.3	6%	—
Bicycles per panorama	0.00 (0.02)	0.00	0.00	0.00	0.00	0.02	57%	0.00
Bike-rack "objects" per 100 units	38.7 (23.7)	0.0	25.8	37.3	51.3	78.5	6%	24.8
Bin "objects" per 100 units	21.1 (11.6)	3.0	15.0	20.3	26.8	42.3	0%	13.4
Bus-stop "objects" per 100 units	5.0 (3.1)	0.7	2.8	4.7	6.8	9.7	3%	3.0
EV-charger "objects" per 100 units	0.26 (0.57)	0.00	0.00	0.00	0.32	1.14	57%	0.15
NAIP tree canopy (%)	23.8 (12.3)	7.5	14.4	21.5	31.1	47.3	0%	—
NAIP vegetation (%)	49.6 (17.0)	18.6	38.6	50.5	62.3	75.5	0%	—
NAIP impervious (%)	49.6 (17.3)	23.1	36.2	48.1	61.0	81.4	0%	—
NAIP roads + other paved (%)	34.1 (10.8)	16.3	26.4	34.9	40.8	49.9	0%	—
NAIP buildings (%)	15.6 (8.3)	6.0	9.6	13.6	19.7	31.8	0%	—

Version B, primary window; 249 usable campuses, 248 with NAIP. Version A: all panoramas and stored shares (road, sidewalk, and building include padding), detections per 100 panoramas. Quoted object names are detector labels; the detections failed screening (Section 5.4).

The measures sort campuses by type and setting (Figure S5; Table S6). Liberal arts colleges had a median green view of 13.5%, against 9.2% for R1 universities, 2.0 vehicles per panorama against 2.8, and 0.03 persons against 0.13. Urban campuses had less canopy (NAIP 19% against 28% for suburban campuses) and more vehicles and persons. The measures are also correlated with each other (Figure 11). Street green view rises with NAIP canopy and vegetation ($\rho \approx 0.6$) and falls with NAIP impervious cover. Vehicles, persons, building share, and NAIP building cover form a second, "urban" cluster, and the bin and EV-charger "objects" correlate with it (ρ up to 0.7 for bins and vehicles), consistent with detectors that respond to busy scenes rather than to the named objects.

Associations between single measures and the outcomes are weak (Table 4; Figure 11, bottom rows). Of 120 measure–outcome pairs, one survives a false-discovery-rate correction at $q < 0.05$: detected bicycles per panorama with QS Sustainability ($\rho = 0.45$). Persons per panorama with the Trans-portation sub-score ($\rho = 0.32$, $q = 0.05$) and the sidewalk-to-road ratio with Transportation ($\rho = 0.26$) point in the expected direction. No greenery measure is associated with the Grounds sub-score ($|\rho| \leqslant 0.18$). The outcomes themselves are distributed very differently (Figure S6; Table S7). STARS overall scores cluster between 60 and 72 points, sub-scores span the full range, QS Sus-tainability scores are strongly right-skewed, and the STARS and QS scores of the same campuses correlate only weakly ($\rho = 0.24$, $n = 68$).

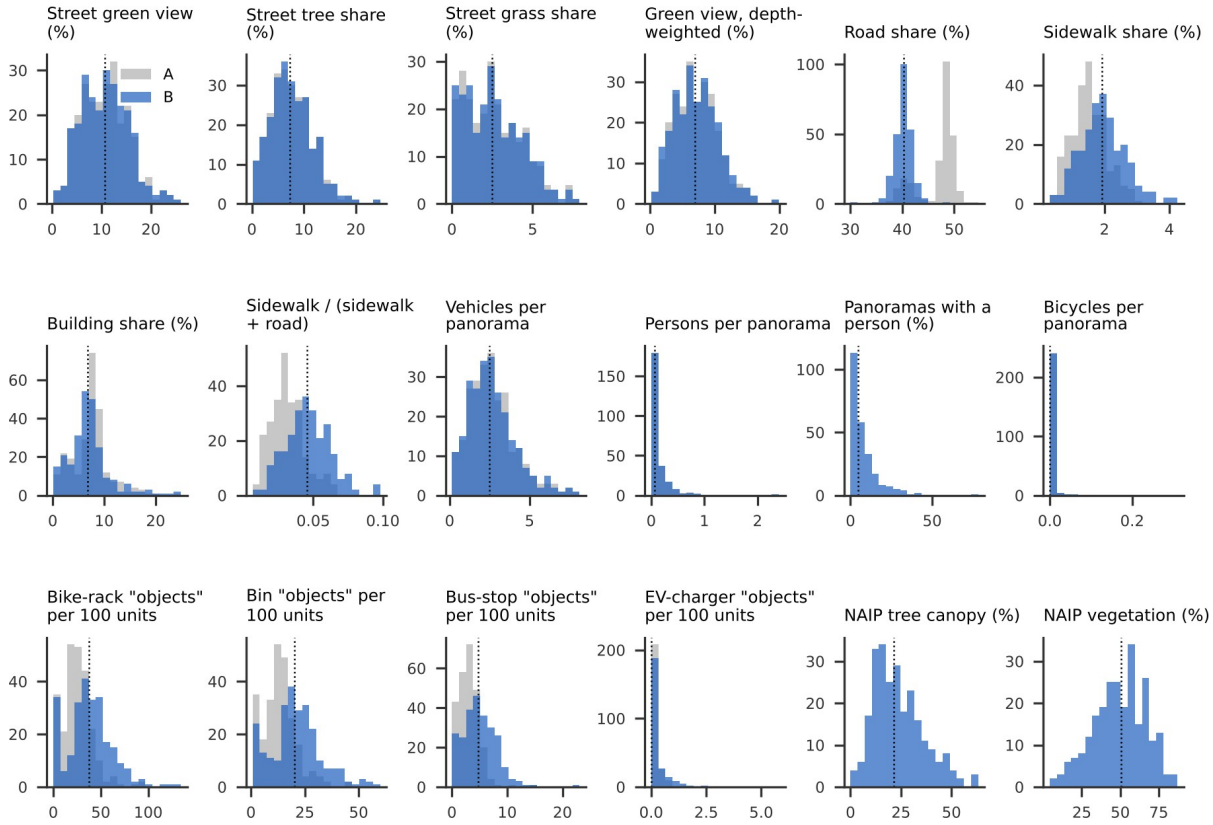

Figure 10: Distributions of campus-level measures in Versions A and B. Gray: Version A; blue: Version B; primary window; dotted line: Version B median. Object measures are per 100 panoramas in A and merged objects per 100 sampling units in B.

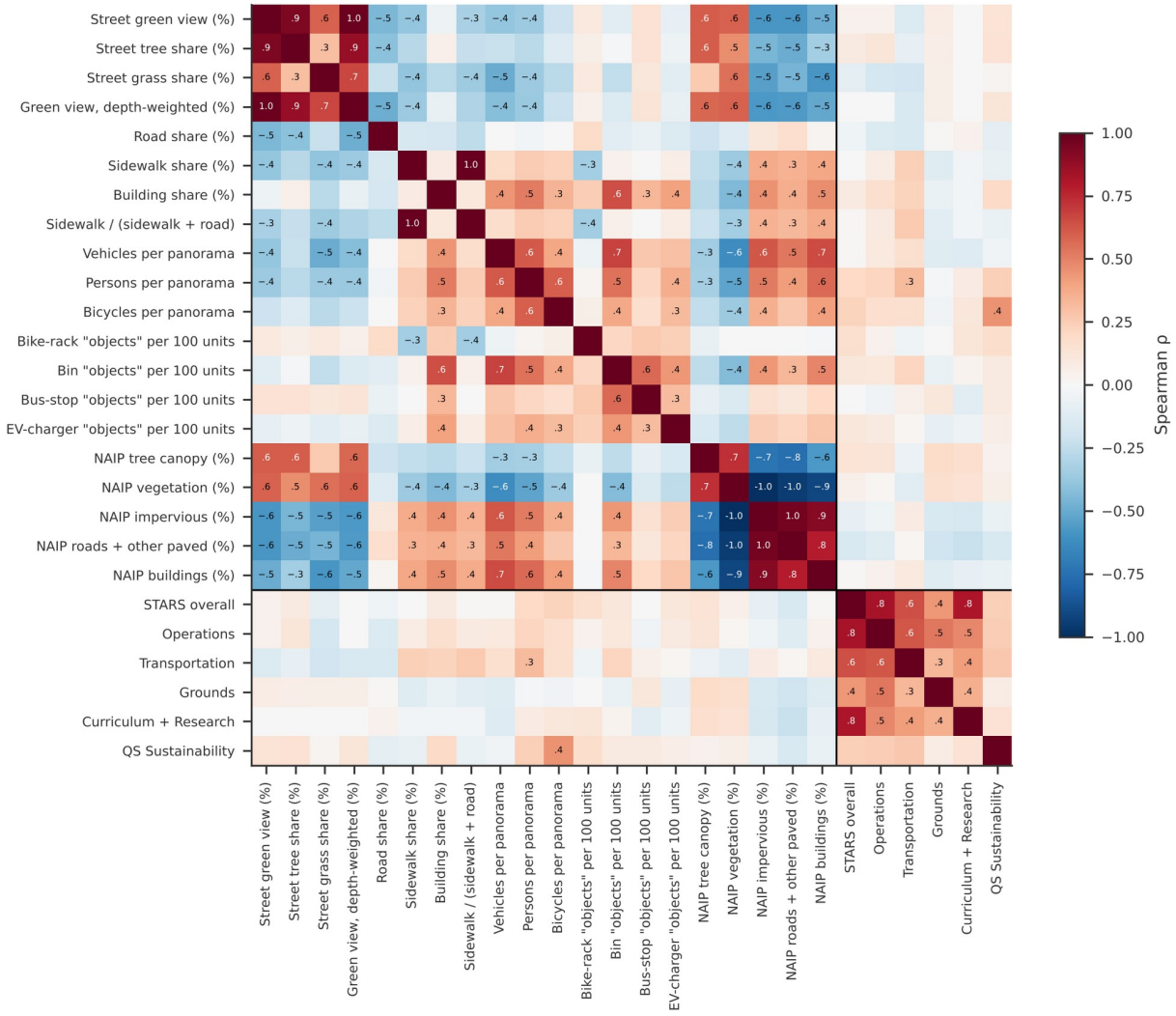

Figure 11: Correlations among measures and outcomes. Spearman ρ, Version B, primary window; values shown where $|\rho| \geqslant 0.3$; lines separate measures from outcomes.

Table 4: Correlations between campus-level measures and outcomes.

Measure	STARS overall	Operations	Transportation	Grounds	Curriculum + Research	QS Sustainability
Street green view (%)	0.04	0.04	−0.13	0.07	0.01	0.13
Street tree share (%)	0.09	0.12	−0.08	0.06	0.00	0.15
Street grass share (%)	−0.10	−0.18	−0.19	0.06	0.00	0.02
Green view, depth-weighted (%)	0.01	0.00	−0.15	0.06	0.00	0.12
Road share (%)	−0.04	−0.14	−0.15	0.01	0.04	−0.10
Sidewalk share (%)	0.00	0.07	0.24	−0.12	−0.03	−0.07
Building share (%)	0.09	0.16	0.22	−0.06	−0.03	0.19
Sidewalk / (sidewalk + road)	0.01	0.09	0.26	−0.10	−0.02	−0.06
Vehicles per panorama	0.01	0.10	0.16	−0.12	−0.13	−0.03
Persons per panorama	0.20	0.22	*0.32*	−0.02	0.08	0.24
Bicycles per panorama	0.22	0.16	0.16	−0.04	0.11	**0.45**
Bike-rack "objects" per 100 units	0.16	0.04	−0.04	0.00	0.12	0.17
Bin "objects" per 100 units	0.11	0.11	0.22	−0.07	0.03	0.09
Bus-stop "objects" per 100 units	0.01	−0.04	0.05	0.11	−0.11	0.09
EV-charger "objects" per 100 units	0.11	0.09	0.02	−0.08	−0.05	0.07
NAIP tree canopy (%)	0.13	0.13	−0.06	0.18	0.17	0.04
NAIP vegetation (%)	0.05	0.02	−0.10	0.16	0.14	0.06
NAIP impervious (%)	−0.08	−0.04	0.08	−0.16	−0.16	−0.07
NAIP roads + other paved (%)	−0.15	−0.12	0.02	−0.20	−0.22	−0.15
NAIP buildings (%)	0.01	0.04	0.10	−0.12	−0.08	−0.06

Spearman ρ, Version B, primary window. n = 128 (STARS overall), 107 (sub-scores), and 113 (QS). Bold: Benjamini–Hochberg $q < 0.05$ over all 120 tests; italic: $q < 0.10$. Correlations with $|\rho| \geq 0.20$ have unadjusted $p < 0.05$.

5.4 Validity of the measures

Street-level green view tracked overhead greenery. Across 248 campuses, Spearman ρ was 0.61 with NAIP tree canopy, 0.59 with NAIP vegetation, and −0.58 with impervious cover (Figure 12a). For some campuses, season matters as much as place. For the 83 campuses whose primary window fell outside May–September, green view in the best leaf-on window was a median 1.31 times higher (Wilcoxon $p < 0.001$; Figure 12b). Across the whole corpus, mean green view falls by about 90% from its peak month to its lowest month in the Midwest and Northeast (Figure S4). Comparing street-level greenery across campuses therefore requires leaf-on windows or a seasonal adjustment.

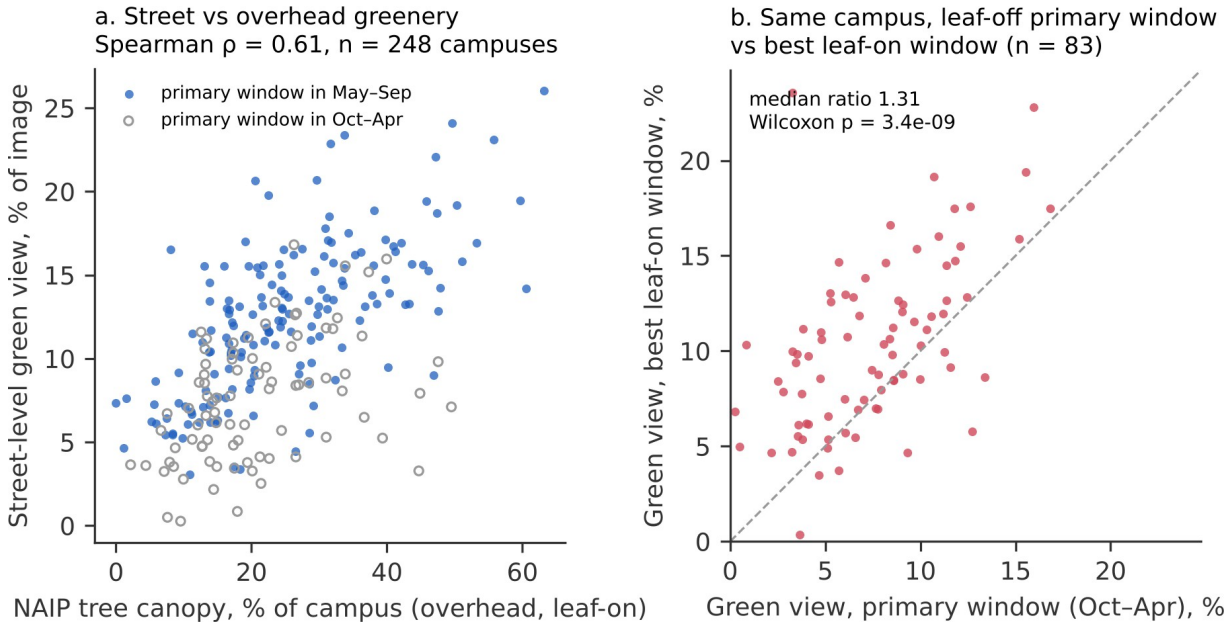

Figure 12: Street greenery, overhead canopy, and season. (a) Street green view (Version B) against NAIP tree canopy across campuses. (b) Green view of the same campus in a leaf-off primary window and in its best leaf-on window.

The special-object detection results are reported in Table 5. At confidence $\geq$ 0.80, bike-rack predic-tions frequently covered entire image slices: 95.2% of the 702,684 boxes assessed had a longest side spanning at least 85% of a 2,048-px slice, with a median size of 1, 871 × 1, 940 px. Inspection showed that these boxes often enclosed road surfaces and sky. In the random screen, 4 of 96 detections were judged correct (4.2%; 95% Wilson interval 1.6–10.2%), and another 4 were uncertain. False posi-tives commonly included roads, cars, windows, doors, signs, and facades. These failures were not apparent from the reported test metrics, which were calculated on curated, non-panoramic images; the bike-rack and garbage-bin test sets also contained no negative images (Table S11). The resulting detections therefore cannot be interpreted as counts of the named objects. We excluded these mea-sures from the index but retained them as image features in the prediction models (Section 5.6). Any predictive contribution must consequently be interpreted with caution, since it may reflect capture year, scene type, or other image characteristics rather than the presence of the intended objects.

Table 5: Special-object detections: resolution artifact and screening.

Object	Fileswithadetection,16,384px	Fileswithadetection,4,096px	Tile-sizedboxes,16,384 px	Screen:correct/uncertainof 24	Precision(95 %CI)	Detector test mAP50
Bike racks	53.8%	0.26%	95.2%	0/0	0%(0–14%)	0.38
Garbage bins	27.3%	4.5%	0.1%	2/1	8%(2–26%)	0.71
Bus stops	6.3%	1.7%	1.8%	2/3	8%(2–26%)	0.48
EV chargers	0.48%	0.007%	0%	0/0	0%(0–14%)	0.19
All four	4.2%	(1.6–10.2%)				

Detections at confidence ≥ 0.80. Shares of files are over all 615,495 files at 16,384 px (including 74,416 blank files; over valid files the shares are 61.2%, 31.0%, 7.2%, and 0.55%) and over 1,878,608 files at 4,096 px. Tile-sized: longest box side ≥ 85% of the 2,048-px slice. Each screened detection was judged once, not by independent labelers. Detector training and test results: Tables S10 and S11.

5.5 Sensitivity to preprocessing

Vegetation measures were nearly identical across versions: campus green-view means differed by 0.01 percentage points on average, with Spearman ρ = 0.998–0.999. Vehicle and person measures were similarly stable ($\rho \geq 0.98$). Road and sidewalk shares were more sensitive (ρ = 0.57 and 0.92, respectively), indicating that preprocessing affected some campus comparisons substantially (Table 6; Figure 13; Table S5).

Object merging removed few repeated detections, with median ratios of 1.00–1.10 detections per merged object. Given the high false-positive rate identified in the screening, neither version provides reliable counts of the intended objects.

5.6 Can imagery reproduce the sustainability scores?

We made a sustained attempt to estimate the scores from the imagery, and this section reports how far it got. No predictor here is derived from a sustainability rating. STARS scores are estimated from imagery and, where stated, institutional covariates, never from QS scores, other STARS scores, or the STARS version, and QS scores are never estimated from STARS.

We started with one row per campus and imagery from the primary window, and the models did poorly. The all-learner average reached R^2 0.02 for the STARS overall score, between −0.09 and −0.02 for the Operations, Transportation, and Grounds sub-scores, and 0.02 for the negative control. The best of the 27 single learners, a degree-2 polynomial with a lasso penalty, reached 0.04 for the overall score. QS Sustainability was the exception (0.24). The choice of learner mattered less than the outcome (Figure 14; Table S2). For every outcome on which any learner had skill, the best family was linear regression, mostly penalized, or, for the STARS overall score, its polynomial expansion. Kernel, spline, nearest-neighbor, tree, and neural-network learners did not improve on

Table 6: Version A versus Version B.

Indicator	VersionAmean(SD)	VersionBmean(SD)	Spearman ρ,AvsB	Rankshift,median/max
Streetgreenview(plain)	10.71%(4.81)	10.72%(4.83)	0.999	2/19
Streettreeshare	7.67%(4.03)	7.68%(4.04)	0.999	1/12
Streetgrassshare	2.62%(1.75)	2.62%(1.74)	0.999	2/13
Streetgreenview(depth-weighted)	7.22%(3.37)	7.24%(3.37)	0.998	2/18
Roadshare	47.03%(3.47)	40.23%(1.74)	0.575	38/195
Sidewalkshare	1.52%(0.54)	1.95%(0.69)	0.918	16/105
Buildingshare	7.43%(3.64)	7.05%(4.18)	0.988	4/58
Vehiclesperpanorama(conf ⩾ 0.50)	2.68(1.48)	2.67(1.48)	0.998	3/15
Personsperpanorama(conf ⩾ 0.50)	0.13(0.21)	0.13(0.21)	0.983	5/63
Bikeracks:Adetectionsper100 panoramas/Bobjectsper100 units(median)	23.6	37.3	0.99	
Garbagebins(same)	13.0	20.3	0.99	
Busstops(same)	2.9	4.7	0.99	

Primary window, 249 usable campuses. Object rows use different denominators (panoramas in A, sampling units in B), so only their rank agreement is comparable. Road, sidewalk, and building shares differ between versions mainly through the padding correction (Section 4.3); with one panorama per unit but stored shares, ρ with Version A is 0.996 for road and sidewalk.

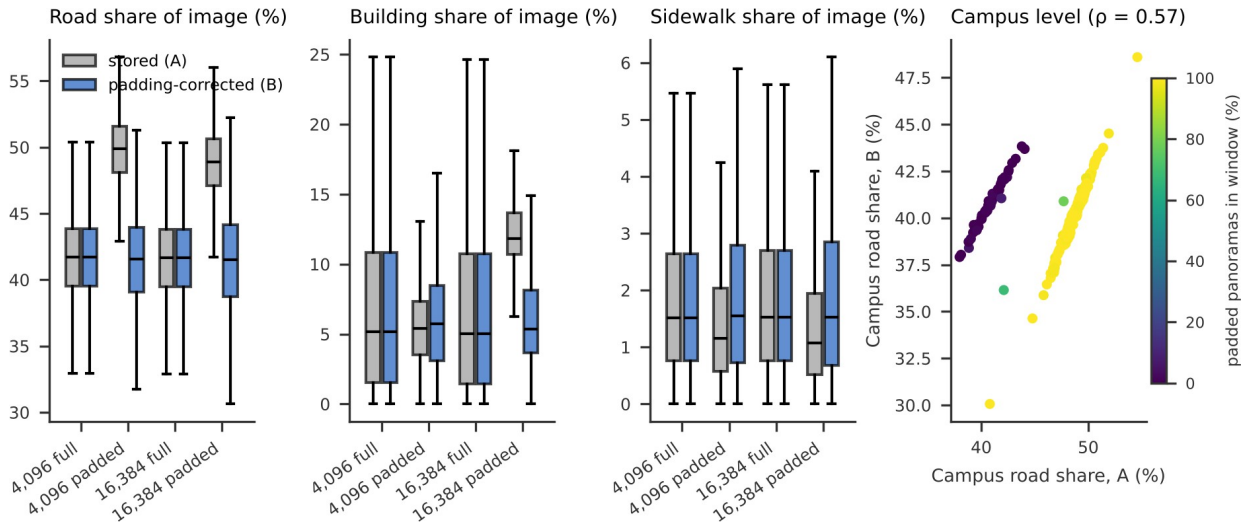

Figure 13: Padding bias in segmentation shares. (a–c) Panorama-level road, building, and sidewalk shares by canvas type, as stored (gray; Version A) and after removing the contribution of the black padding (blue; Version B); boxes show the interquartile range and median, whiskers 1.5 IQR (random sample of 300,000 valid panoramas). (d) Campus road share in Versions A and B, colored by the share of padded panoramas in the window.

them, and gradient boosting was the weakest family for five of the six outcomes, as one would expect with about a hundred campuses and some forty correlated features. Where imagery had any skill, nested model selection and stacking were no better than the plain average of all learners, because choosing among learners on inner folds of 80–100 campuses is itself noisy.

Using every window that covered at least a fifth of a campus's coverable points and at least 20 units improved the estimates (Table 7). Averaging the features of those windows before learning (the multi-window design) raised the STARS overall estimate from 0.02 to 0.08 and Operations from −0.02 to 0.07. Training on each window as a separate row (the window-augmented design) did better still for QS Sustainability (0.20 to 0.27), slightly better for STARS overall (0.08 to 0.10) and Transportation (0.06 to 0.08), and about as well for Operations, Grounds, and the control. From imagery alone, the window-augmented all-learner average reached R^2 0.10 for the STARS overall score, 0.07 for Operations, 0.08 for Transportation, −0.03 for Grounds, and 0.27 for QS Sustainability, against 0.05 for the Curriculum and Research negative control. With whole states held out, these figures were 0.04, 0.05, 0.06, −0.04, 0.26, and 0.01. The best single learners, penalized linear or kernel models wherever imagery had skill, were at most 0.03 above the average. Using only the window closest to each STARS submission (without its capture year, which would carry the submission date) gave R^2 0.08 from imagery for the STARS overall score, −0.01 to 0.01 for the sub-scores, and 0.11 for the control (Table 7; Table S2). Matching the imagery to the report date thus helped the negative control more than the overall score, which again points to what kind of campus is photographed, and when, rather than to visible sustainability practice.

Table 7: Out-of-sample R^2 by approach.

			One row per campus			Window-augmented		
Outcome	*n*	Institutional covariates	Primary window	STARS-matched	Multi-window	Imagery	Image content only	Institutional + imagery
STARS overall	128	0.05	0.02	0.08	0.08	**0.10**	0.08	**0.11**
		(−0.01)	(−0.01)	(0.10)	(0.04)	(0.04)	(0.02)	(0.04)
Operations	107	−0.02	−0.02	−0.01	0.07	**0.07**	0.07	**0.06**
		(0.00)	(0.00)	(−0.01)	(0.04)	(0.05)	(0.07)	(0.05)
Transportation	107	−0.02	−0.02	0.01	0.06	**0.08**	0.08	**0.04**
		(−0.03)	(−0.06)	(0.02)	(0.03)	(0.06)	(0.07)	(0.03)
Grounds	107	0.00	−0.09	−0.01	−0.02	**−0.03**	−0.02	**−0.04**
		(−0.03)	(−0.02)	(0.00)	(−0.03)	(−0.04)	(−0.04)	(−0.04)
Curriculum + Research (control)	107	0.02	0.02	0.11	0.05	**0.05**	0.02	**0.07**
		(−0.01)	(0.01)	(0.10)	(0.06)	(0.01)	(0.00)	(0.03)
QS Sustainability	113	0.37	0.24	n/a	0.20	**0.27**	0.20	**0.39**
		(0.38)	(0.22)		(0.13)	(0.26)	(0.18)	(0.40)

All-learner average, mean of three repeats of 5-fold cross-validation; whole states held out in parentheses; imagery only unless stated. The all-learner average is the mean prediction of every learner (27 for one row per campus, nine for the window-augmented models) and involves no model selection. The same campuses enter every column of a row, except the STARS-matched window (one campus fewer). STARS-matched window: the window closest to the STARS submission, used without its capture year; not applicable to QS. Institutional covariates: log enrollment, Carnegie class, locale, and census region. Multi-window average: one row per campus, features averaged over its windows. Window-augmented: one row per window; the campus prediction is the mean over its windows. Image content only: without each window's capture metadata (coverage, number of units, capture year, leaf-on flag, and share of high-resolution files). No predictor is derived from any sustainability rating. Every learner and pipeline is in Table S2.

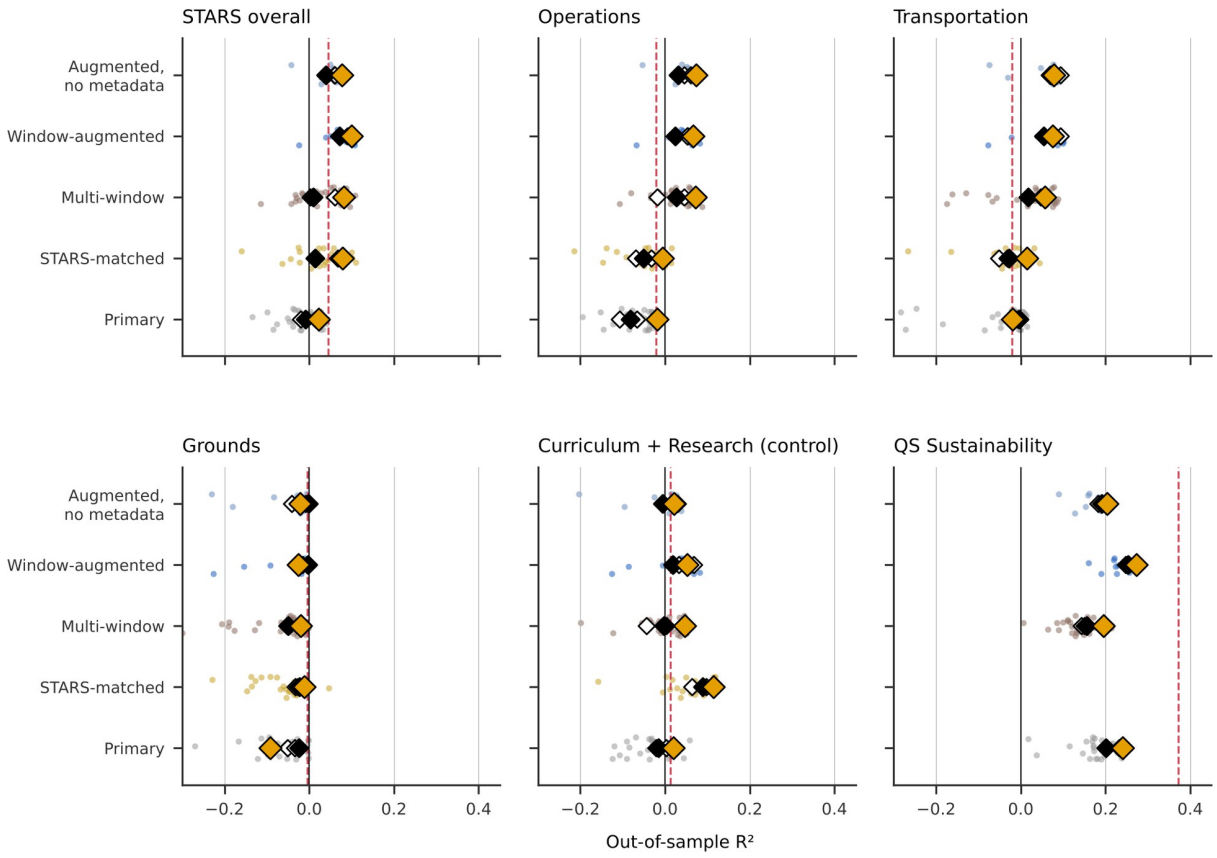

Figure 14: Out-of-sample R^2 by outcome and approach. Imagery only; mean of three repeats of 5-fold cross-validation. Dots: single learners. Diamonds: pipelines fixed in advance (orange: average of all learners, as in Table 7; white: nested selection and top-five average; black: stacking). Red dashed line: institutional covariates alone. The STARS-matched window applies to STARS outcomes only.

Permuting one block of features at a time across the held-out campuses of the window-augmented models shows where the accuracy came from (Figure 15b). Greenery contributed nothing to any outcome, and neither did road, sidewalk, and building shares or the special objects (drops in R^2 of at most 0.01). People, bicycles, and vehicles in view were the largest block for Operations (0.07) and Transportation (0.12). For QS Sustainability they tied with capture metadata (0.11 each), and NAIP land cover added 0.04. For the STARS overall score, the same three blocks contributed about equally (0.04, 0.04, and 0.02). For the negative control, the people and vehicles block contributed nothing (−0.01) and capture metadata the most (0.07). Removing the capture metadata altogether (Table 7, image content only) left Operations at 0.07 and Transportation at 0.08, lowered the STARS overall estimate to 0.08 and QS to 0.20, and reduced the control to 0.02 (0.00 with states held out).

Reassigning whole campuses' imagery to other campuses at random gave null R^2 values centered near or below zero (Table 8; Figure 15c). Against this null, imagery alone was significant at the 5% level for STARS overall, Transportation, the control, and QS Sustainability ($p \leqslant 0.039$) but not for Operations ($p = 0.06$) or Grounds ($p = 0.37$). That the negative control passes this test shows that imagery can be informative about a campus without seeing what a score measures. Added to institutional covariates, imagery raised R^2 from 0.05 to 0.11 for the STARS overall score, from −0.02 to 0.06 for Operations, from −0.02 to 0.04 for Transportation, and from 0.37 to 0.39 for QS. It also raised R^2 from 0.02 to 0.07 for the negative control, a gain (0.055) almost as large as for the overall score (0.058). When imagery was reassigned while each campus kept its institutional covariates, the gain was larger than chance for STARS overall, Transportation, and QS Sustainability, but not for Operations ($p = 0.10$), Grounds ($p = 0.67$), or the control ($p = 0.19$). In sample, ordinary least squares on the ten street-level measures explained between 6% and 16% of the variance (adjusted R^2 at most 0.08; Table S3). With all imagery features the in-sample R^2 rose to 0.39–0.48, but the adjusted R^2 stayed at or below 0.13, and the ten measures did not significantly improve on institutional covariates for any outcome (partial F-test $p \geqslant 0.08$).

Table 8: Permutation tests and gains over institutional covariates.

	Imagery alone			Added to institutional covariates					
Outcome	R^2	Null 95th pct	p	Institutional + imagery R^2	Institutional R^2	Gain	Gain, states out	Null 95th pct	p
STARS overall	0.10	0.02	0.020	0.05	0.11	0.06	0.05	0.08	0.048
Operations	0.07	0.04	0.059	−0.02	0.06	0.08	0.05	0.04	0.095
Transportation	0.08	0.04	0.020	−0.02	0.04	0.06	0.05	0.05	0.048
Grounds	−0.03	0.02	0.373	0.00	−0.04	−0.04	−0.01	0.00	0.667
Curriculum + Research (control)	0.05	0.03	0.039	0.02	0.07	0.06	0.05	0.08	0.190
QS Sustainability	0.27	0.03	0.020	0.37	0.39	0.02	0.02	0.35	0.048

Window-augmented models. Imagery R^2, institutional R^2, and the gain: all-learner average, three repeats of 5-fold cross-validation (institutional covariates alone: one row per campus, same campuses). Nulls: whole campuses' imagery (all windows and NAIP shares) reassigned at random to other campuses, with scores (and, for the second test, institutional covariates) kept in place; 50 permutations for imagery alone and 20 with institutional covariates; three learners (ridge, kernel ridge, and extremely randomized trees) and their average refitted on the first random split. $p = (k + 1) / (K + 1)$, where k of K permutations reached the R^2 of the same three-learner average on the observed data. Because that observed three-learner R^2 can differ from the all-learner R^2 shown, an R^2 above the null 95th percentile can still have p above 0.05, and the reverse.

The two measurement versions estimated the scores about equally poorly. Restricted to the ten

measures of the primary window, out-of-sample R^2 for Version B minus Version A ranged from −0.03 to +0.04 across the six outcomes, and in-sample R^2 was 0.05–0.16 for A and 0.06–0.16 for B (Tables S2, S3, and S3b). The padding correction changed which campuses rank high on road share (Section 5.5) but not how much the measures say about the ratings.

Because the first search left so much unexplained, we widened it, and the wider search found no more accuracy (Table 9). On identical splits, 33 learners each tuned over wide random ranges (about 5,500 settings per run) matched the earlier search on the same features. The all-learner average differed by −0.01 to +0.02 on the random split and by −0.01 to +0.03 with whole states held out, and for the STARS overall score in the window-augmented design it reached 0.12, the same as before (0.04 against 0.04 with states held out; Table S2). With the same learners and one row per campus, replacing the six NAIP shares with the full overhead block lowered every STARS estimate with whole states held out, by 0.01 to 0.08 (STARS overall: 0.05 to 0.00). On the random split it lowered the overall score, Operations, and Grounds but raised Transportation (0.07 to 0.09) and the control (0.01 to 0.04), gains that did not survive holding states out. Dropping the overhead image metadata left the STARS overall estimate with states held out at 0.00. In the window-augmented design, where the comparison also widens the search, the overhead block lowered the STARS overall, Operations, and Transportation estimates with states held out (by 0.04 to 0.11) and changed Grounds, the control, and QS Sustainability by at most 0.02. QS Sustainability was the only outcome that the overhead block helped in both splits with one row per campus (0.28 to 0.31; 0.13 to 0.19 with states held out), which fits its link to campus size (ρ = 0.30 with campus area). With institutional covariates, the overhead block reached 0.40 per campus (0.30 with states held out) and 0.41 in the window-augmented design (0.37). Both are below the earlier window-augmented model (0.43; 0.40), and with states held out they are also below institutional covariates alone (0.37; 0.38). The best single learner in any wide-search configuration reached 0.16 for the STARS overall score (partial least squares, window-augmented). Because that is the largest of 194 cross-validated values on one split, it is an upper bound rather than an estimate.

Table 9: Wide search versus earlier search on identical splits.

		Imagery, one row per campus				Imagery, window-augmented		With institutional covariates			
Outcome	*n*	Earlier search	Wide search	Wide + overhead	Wide + overhead, no metadata	Earlier search	Wide + overhead	Covariates only	Earlier, augmented	Wide + overhead, per campus	Wide + overhead, augmented
STARS overall	128	0.10	0.10	0.08	0.09	0.12	0.08	0.06	0.13	0.08	—
		(0.04)	(0.05)	(0.00)	(0.00)	(0.04)	(−0.03)	(−0.01)	(0.04)	(−0.01)	
Operations	107	0.05	0.07	0.04	0.06	0.08	−0.02	−0.02	0.06	0.03	—
		(0.04)	(0.07)	(−0.02)	(−0.01)	(0.05)	(−0.05)	(0.00)	(0.05)	(−0.01)	
Transportation	107	0.07	0.07	0.09	0.09	0.08	0.08	0.01	0.05	0.08	—
		(0.03)	(0.03)	(0.01)	(0.01)	(0.06)	(0.02)	(−0.03)	(0.03)	(0.01)	
Grounds	107	0.02	0.02	0.00	0.01	−0.03	−0.02	−0.02	−0.05	0.00	—
		(−0.03)	(0.00)	(−0.02)	(−0.02)	(−0.04)	(−0.03)	(−0.03)	(−0.04)	(−0.01)	
Curriculum + Research (control)	107	0.02	0.01	0.04	0.06	0.03	0.04	0.03	0.06	0.04	—
		(0.06)	(0.09)	(0.05)	(0.06)	(0.01)	(0.03)	(−0.01)	(0.03)	(0.05)	
QS Sustainability	113	0.27	0.28	0.31	0.31	0.32	0.33	0.37	0.43	0.40	0.41
		(0.13)	(0.13)	(0.19)	(0.19)	(0.26)	(0.25)	(0.38)	(0.40)	(0.30)	(0.37)

Out-of-sample R^2 of the all-learner average on the first random 5-fold split; whole states held out in parentheses. Earlier search: the 27 learners (nine window-augmented) of Table 7. Wide search: 33 learners (31 window-augmented), each tuned by random search inside every training fold (Section 4.5; Table S14). All columns use the same campuses and the same outer folds: the first of the three random 5-fold splits of Table 7 and the state-grouped split. Table 7 averages three random splits, so its values differ. Overhead: the full overhead block of 38 features (Section 3.3) in place of the six NAIP shares; no metadata: without mean band brightness, classification confidence, and the number of acquisitions. Institutional covariates only (log enrollment, Carnegie class, locale, and census region): earlier search, one row per campus. —: not run. No predictor is derived from any sustainability rating. Every learner and pipeline is in Table S2.

Even the best of these models leaves most of the variation unexplained. For the STARS overall score, the out-of-fold predictions of the five-learner window-augmented average on the first random split (R^2 0.12) had a root-mean-square error of 10.2 points, against 10.9 for predicting every campus at its training-fold mean. Their Spearman correlation with the observed scores was 0.30, and they placed 37% of campuses in the correct tercile (chance: 33%). For QS Sustainability the corresponding figures were 26.1 against 31.5, ρ = 0.61, and 52%; for Transportation they were 20.8 against 21.8, ρ = 0.22, and 38% (Figure 15a). The spread of the predictions was only 26% of the spread of the observed STARS scores, so the models separate broad groups of campuses but cannot place individual campuses on the rating scales. Despite the breadth of the search, then, we were not able to reproduce the sustainability scores from the imagery.

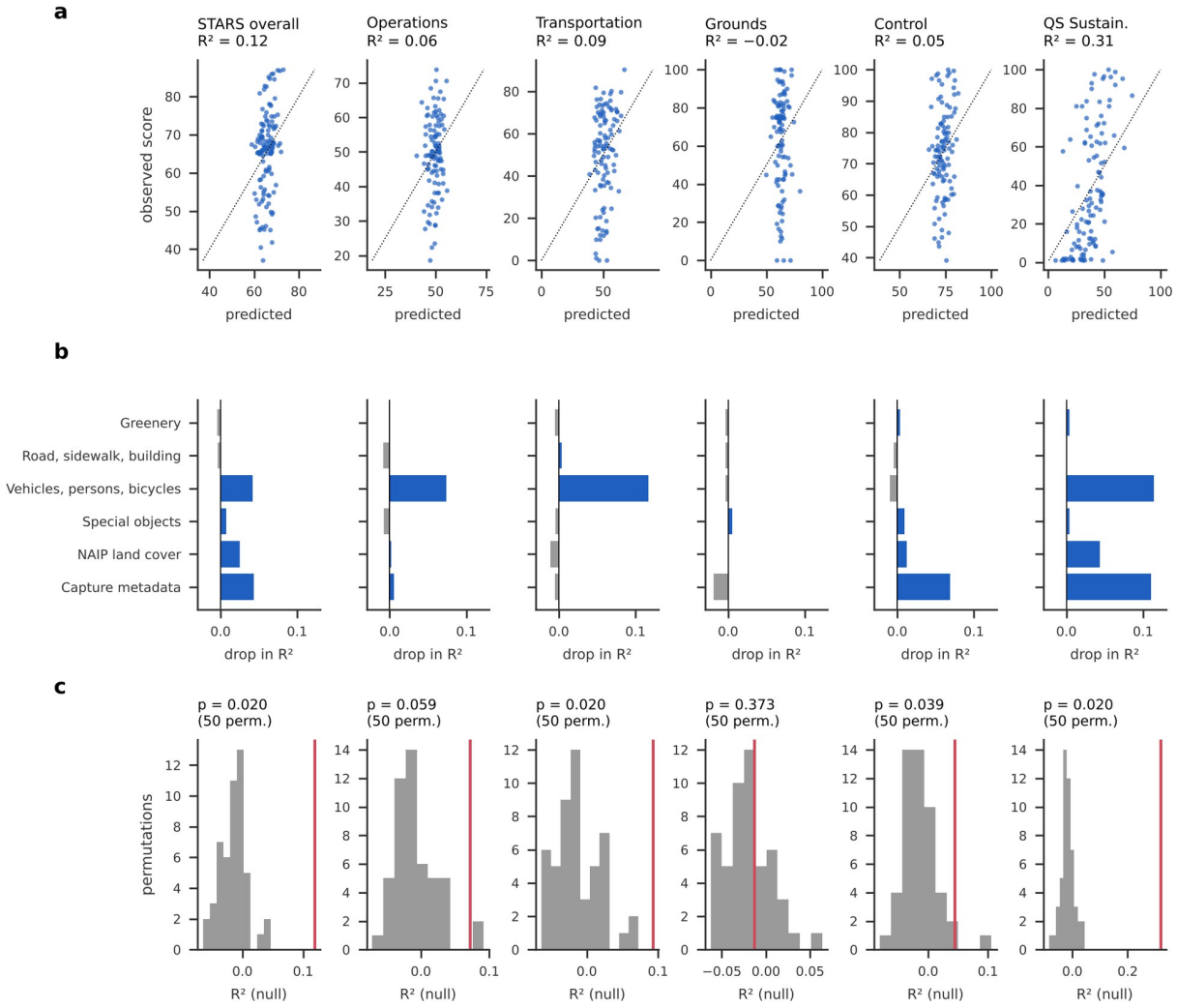

Figure 15: Window-augmented models from imagery alone. First random 5-fold split. (a) Observed against out-of-fold predicted scores (average of five learners). (b) Drop in R^2 when one block of features is permuted across the held-out campuses (mean of 10 permutations per fold). (c) R^2 when whole campuses' imagery is reassigned at random (gray) and observed R^2 (red), average of three learners.

5.7 The Visible Campus Environment Index: A new ranking

With all three dimensions in place, the VCEI ranges from 0.05 to 0.90 across 249 campuses (median 0.42, IQR 0.30–0.55; Version B, primary window; Figure 16; Table S16). The dimensions measure different things. Greenery and low car dominance are strongly correlated (Spearman ρ = 0.78), because green campuses have less paving and fewer vehicles in view. Pedestrian space, by contrast, is negatively correlated with greenery (ρ = −0.33) and almost unrelated to car dominance (ρ = −0.14), since the campuses with the most sidewalk in view are dense urban ones. The ordering follows campus setting, with median index values of 0.59 for rural, 0.48 for town, 0.44 for suburban, and 0.41 for urban campuses. The three highest-ranked campuses are residential colleges (Wellesley, Denison, and Agnes Scott). The three lowest are the U.S. Air Force Academy; the University of California, Merced, a young campus on open land; and Columbia University in Manhattan (Figure 17).

The full index is sensitive to measurement choices. Version A and Version B indices agree only moderately (ρ = 0.84), because the padding correction changes the road and sidewalk indicators (Section 5.5). Primary and leaf-on windows give ρ = 0.86, and geometric and arithmetic aggregation give ρ = 0.85, because the dimensions disagree and the geometric mean penalizes a low pedestrian score. Varying the weights, dropping indicators, and switching the aggregation rule produced 90% rank intervals with a median width of 89 places among 249 (IQR 58–142; Figure 16a). The index separates the clearest cases at the two ends but cannot order most campuses.

The index was only weakly associated with sustainability scores: Spearman ρ = 0.15 with the STARS overall score, 0.19 with Operations, 0.13 with Transportation, 0.02 with Grounds, 0.11 with QS, and 0.14 with the Curriculum and Research negative control. We report these associations as context, not validation. The index was not built to reproduce these scores, and the negative control is about as strongly associated with it as the operational sub-scores are.

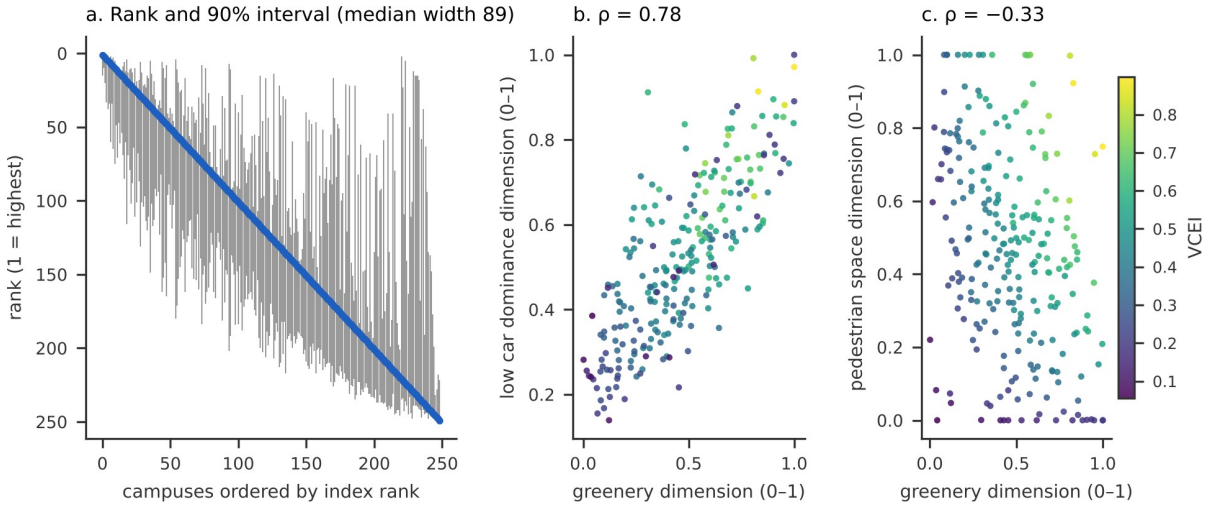

Figure 16: Visible Campus Environment Index. (a) Rank of each campus with its 90% rank interval from 500 draws that vary dimension weights, drop one indicator, and switch the aggregation rule. (b) Greenery against low car dominance. (c) Greenery against pedestrian space. Color shows the index value.

Three highest

Wellesley College
VCEI 0.90, rank 1 of 249



Sep 2012 · dimensions: greenery 1.00
pedestrian space 0.75 · low car dominance 0.97

Denison Univ.
VCEI 0.89, rank 2 of 249



Jul 2007 · dimensions: greenery 0.83
pedestrian space 0.92 · low car dominance 0.91

Agnes Scott College
VCEI 0.85, rank 3 of 249



Jul 2014 · dimensions: greenery 0.95
pedestrian space 0.73 · low car dominance 0.88

Three lowest

United States Air Force Academy
VCEI 0.05, rank 249 of 249



Aug 2015 · dimensions: greenery 0.04
pedestrian space 0.00 · low car dominance 0.39

Univ. of California-Merced
VCEI 0.09, rank 248 of 249



Apr 2019 · dimensions: greenery 0.00
pedestrian space 0.22 · low car dominance 0.28

Columbia Univ.
VCEI 0.09, rank 247 of 249



Apr 2009 · dimensions: greenery 0.04
pedestrian space 0.08 · low car dominance 0.24

Figure 17: Highest- and lowest-ranked campuses on the index. One randomly drawn panorama from the primary window of each of the three highest- and three lowest-ranked campuses, with its dimension scores.

6 Discussion

6.1 What street-level imagery measures on a campus

Street-level imagery measured campus greenery in a way that agrees with an independent overhead source (ρ = 0.61 with NAIP canopy). That is stronger than the point-level correlations of at most 0.50 reported by Larkin and Hystad (2019), as one would expect when many views are aggregated over an area (Stubbings et al., 2019). The agreement depends on season, though. The same campus shows a median 31% more green view in its best leaf-on window than in a leaf-off one, and a third of primary windows fell outside May–September. Any campus comparison of street greenery therefore needs leaf-on windows or an explicit seasonal adjustment, consistent with the large seasonal bias reported across cities (Zhao et al., 2025).

The temporal structure of the imagery is the binding constraint. Two-month windows without chaining cover most of a campus (the best window reaches a median 78% of coverable points). But the share of campuses with a well-covered window peaked in 2017–2019, fell to 11–18% a year in 2020–2023, and nearly vanished in 2024–2025. Annual campus measurement is supported for only about one campus-year in five, and the imagery that covers campuses best was taken a median 9.4 years before the STARS reports it would be compared with. A study that reports annual image-based campus indices, or that treats a decade-old streetscape as concurrent with a recent rating, is measuring something other than what it claims.

6.2 Why the ratings could not be reproduced

We were not able to reproduce the sustainability scores from imagery. Across eight families of learners, flexible models did no better than penalized linear ones, choosing one learner or stacking several did no better than averaging them all, and the largest improvement came from using more of the imagery (every well-covered window) rather than from a more flexible model. A wider search of 33 learners, each tuned at random over wide ranges, reproduced the earlier search on identical splits (STARS overall 0.10 against 0.10). With the same learners, replacing six overhead land-cover shares with 38 overhead features lowered the STARS overall estimate with whole states held out (0.05 to 0.00); many overhead features differ by state, since NAIP is flown state by state. Even at its best, imagery alone explained at most about a tenth of the variance in STARS overall scores (0.10 for the best design in random folds and 0.10 for the best design with whole states held out), 0.07–0.08 for Operations and Transportation, and nothing for Grounds, the credit most directly about visible landscape. The negative control was estimated about as well as Operations and Transportation (0.05), largely from when and how its streets were photographed, and with imagery matched to the date of the STARS report it was estimated better than the overall score (0.11 against 0.08).

Added to institutional covariates, imagery raised cross-validated R^2 by about as much for the control (0.06) as for the overall score (0.06). The permutation test did separate the two (p = 0.048 for the overall score and for Transportation, against 0.19 for the control), so the evidence that imagery adds something specific to the ratings is weak rather than absent. A camera has no reason to see a

curriculum, so much of what imagery adds most plausibly reflects what kind of institution a campus is (its resources, residential character, and setting, and how often its streets were photographed) beyond what enrollment, Carnegie class, locale, and region capture. The block that carried most of the signal for Operations and Transportation was people, bicycles, and vehicles in view. That fits what those credits reward (walking, cycling, and transit), but it is also simply what a large residential campus looks like. QS Sustainability scores were explained mainly by institution size and type (R^2 0.37 from institutional covariates), and imagery added 0.02. The fit was weak in the classical sense as well: least squares on the ten street-level measures explained 6–16% of the variance in sample, and for no outcome did they significantly improve a least-squares model of institutional covariates.

We cannot rule out that other imagery, validated detectors, or a much larger sample would do better. Within what we could test, however, the most likely explanation is that the visible campus environment and institutional sustainability ratings are different constructs. STARS scores combine academic, engagement, and governance credits with operational ones, and the operational credits concern energy, purchasing, waste, and travel behavior that a street car does not photograph (AASHE, 2024b). A greener campus is not thereby a better-rated one, and a well-rated institution may sit on a dense urban site. This reading is consistent with critiques that assessment tools rely on internal proxy data (Findler et al., 2019) and embed implicit norms of what a sustainable university is (Fischer et al., 2015), with evidence that STARS reporting improves with practice (Minutolo et al., 2021), and with evidence that STARS ratings track institution size and type and do not reliably predict emissions (Aldogan Eklund, 2026; Doocy et al., 2021). It also argues against image-based campus "sustainability" scores calibrated to ratings. With about a hundred campuses, small cross-validated gains are easy to obtain and easy to overread (Varoquaux, 2018), and our negative control shows that such gains can reflect campus type rather than the construct of interest.

We therefore offer the VCEI as a measure of visible campus conditions in its own right. Greenery and low car dominance move together across campuses (ρ = 0.78), while pedestrian space runs against greenery (ρ = −0.33), so the index mainly separates leafy, low-traffic campuses from dense urban ones. Its 90% rank intervals have a median width of 89 places among 249 campuses, so it supports statements about broad groups of campuses rather than league-table positions.

6.3 Lessons for street-level measurement

Four lessons apply beyond campuses. The first is to audit the image files before analyzing them. Half of this corpus carried a duplicated 27.7° sector and black padding, and 3% of files were blank; none of this showed in the processing logs.

The second is to know what each model was run on. Thinning to one panorama per 20 m unit hardly mattered at the campus level (rank correlations of at least 0.98 for greenery, vehicles, and persons), and merging repeated sightings changed object counts by about 10% while leaving rankings unchanged. But the segmentation model for road, sidewalk, and building had been run on padded canvases and labeled the black padding as road and building. Because padded files dominate earlier capture years, the stored road shares ranked campuses partly by file format (ρ = 0.57 with the

corrected shares). Any pipeline that mixes canvas formats should check per-class outputs on blank and padded inputs before aggregating.

The third is that object detectors trained on small curated image sets did not transfer to tiled panoramas. A random screen found 4 of 96 high-confidence detections correct, and 95% of high-resolution bike-rack boxes were nearly tile-sized (longest side at least 85% of the 2,048-px tile). Test metrics computed on sets without negative images give no warning of this (Supplementary Note S2), so in-situ validation on the target imagery should come before detections are used as measurements.

The fourth concerns prediction with about a hundred units. Repeated and spatially grouped cross-validation, permutation references, and a negative-control outcome together kept us from reading small cross-validated gains as evidence about sustainability, and reporting the average of all learners rather than the best of them kept the accuracy estimates honest (Roberts et al., 2017; Kapoor & Narayanan, 2023).

6.4 Limitations

The outcomes are self-reported, voluntary ratings of a selected set of institutions (STARS participants and top-ranked campuses), and they mix STARS versions 2.2 and 3.0; sub-scores were unavailable for 21 expired reports. The imagery is a proprietary record collected for other purposes, sparse in recent years, and dated only to the month. The vegetation model may undercount dense canopy (Figures 6 and 7), and the bias of equirectangular pixel shares has not been measured on masks. Capture-month climate was not available, so season is represented by the calendar month.

The padding correction for road, sidewalk, and building shares rests on class rates estimated by comparing padded and full canvases photographed at the same places on different dates. It removes the average bias but not panorama-level variation in how padding is labeled, which only re-run masks can measure. The detection screening judged each detection once rather than by independent labelers. Within-campus variation (Figure 9) enters only through campus means, percentiles, and repeated windows; it is not modeled spatially.

7 Conclusion

We set out to reproduce campus sustainability ratings from what street-level and overhead imagery show, and we could not. Large street-level image collections can describe campus environments, greenery above all, provided the images are audited, grouped into short leaf-on observation windows, and interpreted with their dates in view. But across 27 learners and every window design we tried, the average of all learners explained at most about 10% of the variation in STARS overall scores out of sample (mean of three random splits). A wider search of 33 tuned learners did no better on identical splits, and neither did a fuller set of 38 overhead land-cover features. The models estimated the Operations and Transportation sub-scores about as well as a negative-control sub-score, and their small gains over institutional covariates, although they passed a permutation test for the overall score, were matched in size by the control. They missed the Grounds sub-score

entirely and added little to institutional covariates for QS scores. Visible campus conditions are worth measuring in their own right, with explicit uncertainty, but on our evidence they are not a proxy for institutional sustainability ratings.

Data and code availability

Panorama images cannot be redistributed. We will share the panorama identifiers with their campus links, locations, and capture months; per-panorama vegetation, road, sidewalk, and building shares (stored and padding-corrected) and detection counts; and the estimated padding rates. At the campus level, we will share the indicators for Versions A and B in every window, the wide feature matrix and the window-level feature file used in the model search, and the full overhead land-cover block. For the models, we will share every learner's cross-validated R^2 by outcome, window, and feature set, with out-of-fold predictions, permutation nulls, feature-block importances, in-sample least-squares fits, and the setting chosen in every training fold of the wide search. We will also share the STARS sub-scores with the report URL each was taken from, the index with its goalposts, and all analysis code with a run order.

References

Adenle, Y. A., Chan, E. H. W., Sun, Y., & Chau, C. K. (2020). Exploring the coverage of environmental-dimension indicators in existing campus sustainability appraisal tools. *Environmental and Sustainability Indicators*, *8*, 100057. https://doi.org/10.1016/j.indic.2020.100057

Akyon, F. C., Altinuc, S. O., & Temizel, A. (2022). Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 966–970). https://doi.org/10.1109/ICIP46576.2022.9897990

Alberti, C., Meoli, M., & Paleari, S. (2026). University sustainability rankings and sustainability performance: Assessing the relationship. *Sustainability Accounting, Management and Policy Journal*, *17* (7), 163–197. https://doi.org/10.1108/SAMPJ-04-2025-0598

Aldogan Eklund, M. (2026). Green performance in higher education: How institutional and contextual factors shape sustainability ratings. *International Journal of Sustainability in Higher Education*, 1–19. https://doi.org/10.1108/IJSHE-01-2025-0033

Alghamdi, N., den Heijer, A., & de Jonge, H. (2017). Assessment tools' indicators for sustainability in universities: An analytical overview. *International Journal of Sustainability in Higher Education*, *18* (1), 84–115. https://doi.org/10.1108/IJSHE-04-2015-0071

Association for the Advancement of Sustainability in Higher Education [AASHE]. (2019). STARS 2.2 Technical Manual. *AASHE*. https://stars.aashe.org/wp-content/uploads/2019/07/STARS-2.2-Techn ical-Manual.pdf

Association for the Advancement of Sustainability in Higher Education [AASHE]. (2024a). STARS 3.0. *Web page*. https://stars.aashe.org/resources-support/stars-3-0/

Association for the Advancement of Sustainability in Higher Education [AASHE]. (2024b). STARS Technical Manual, Version 3.0. *AASHE*. https://stars.aashe.org/wp-content/uploads/2024/05/STARS-Tec

hnical-Manual-v3.0.pdf

Bai, Y., Cao, M., Wang, R., Liu, Y., & Wang, S. (2022). How street greenery facilitates active travel for university students. *Journal of Transport & Health*, *26*, 101393. https://doi.org/10.1016/j.jth.2022.101393

Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, *80*, 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305. https://jmlr.org/papers/v13/bergstra12a.html

Berzosa, A., Bernaldo, M. O., & Fernández-Sanchez, G. (2017). Sustainability assessment tools for higher education: An empirical comparative analysis. *Journal of Cleaner Production*, *161*, 812–820. https://doi.org/10.1016/j.jclepro.2017.05.194

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, *215*, 104217. https://doi.org/10.1016/j.landurbplan.2021.104217

Biljecki, F., Zhao, T., Liang, X., & Hou, Y. (2023). Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives. *International Journal of Applied Earth Observation and Geoinformation*, *122*, 103385. https://doi.org/10.1016/j.jag.2023.103385

Boiocchi, R., Ragazzi, M., Torretta, V., & Rada, E. C. (2023). Critical analysis of the GreenMetric World University Ranking system: The issue of comparability. *Sustainability*, *15* (2), 1343. https://doi.org/10.3390/su15021343

Breiman, L. (1996). Stacked regressions. *Machine Learning*, *24* (1), 49–64. https://doi.org/10.1007/BF00117832

Bullock, G., & Wilder, N. (2016). The comprehensiveness of competing higher education sustainability assessments. *International Journal of Sustainability in Higher Education*, *17* (3), 282–304. https://do i.org/10.1108/IJSHE-05-2014-0078

Cai, B. Y., Li, X., Seiferling, I., & Ratti, C. (2018). Treepedia 2.0: Applying deep learning for large-scale quantification of urban tree cover. In *2018 IEEE International Congress on Big Data (BigData Congress)* (pp. 49–56). https://doi.org/10.1109/BigDataCongress.2018.00014

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11* (70), 2079–2107. https://jmlr.org/papers/v11/cawley10a.html

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Computer Vision – ECCV 2018* (pp. 833–851). https://doi.org/10.1007/978-3-030-01234-2_49

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Chen, X., Meng, Q., Hu, D., Zhang, L., & Yang, J. (2019). Evaluating greenery around streets using Baidu panoramic street view images and the panoramic green view index. *Forests*, *10* (12), 1109. https://doi.org/10.3390/f10121109

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1280–1289). https://doi.org/10.1109/CVPR52688.2022.00135

Colorado State University. (n.d.). Rankings at Colorado State University. *Admissions*. https://admissions.colostate.edu/about-csu/rankings/

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The Cityscapes dataset for semantic urban scene understanding. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3213–3223). https://doi.org/10.1109/CVPR.2016.350

Cortese, A. D. (2003). The critical role of higher education in creating a sustainable future. *Planning for Higher Education*, *31* (3), 15–22. https://www.scup.org/resource/the-critical-role-of-higher-education-in-creating-a-sustainable-future/

Curtis, J. W., Curtis, A., Mapes, J., Szell, A. B., & Cinderich, A. (2013). Using Google Street View for systematic observation of the built environment: Analysis of spatio-temporal instability of imagery dates. *International Journal of Health Geographics*, *12* (1), 53. https://doi.org/10.1186/1476-072X-12-53

Dietterich, T. G., Lathrop, R. H., & Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, *89* (1–2), 31–71. https://doi.org/10.1016/S0004-3702(96)00034-3

Dileep, R., Patil, M., Mahesh, T., Deshmukh, A., Liu, S., Narayanan, A., & McMahan, L. (2024). Predicting college retention rates from Google Street View images of campuses. *Journal of Emerging Investigators*, *7*. https://doi.org/10.59720/23-112

Doocy, L. E., Zarmehr, A., & Kider, J. T., Jr. (2021). A critical review of the effectiveness of the Sustainability Tracking, Assessment & Rating System (STARS) framework on campus sustainability. In *Proceedings of Building Simulation 2021: 17th Conference of IBPSA* (pp. 629–635). https://doi.org/10.26868/25222708.2021.31030

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32* (2), 407–499. https://doi.org/10.1214/009053604000000067

Fan, Z., Feng, C.-C., & Biljecki, F. (2025). Coverage and bias of street view imagery in mapping the urban environment. *Computers, Environment and Urban Systems*, *117*, 102253. https://doi.org/10.1016/j.compenvurbsys.2025.102253

Findler, F., Schönherr, N., Lozano, R., & Stacherl, B. (2019). Assessing the impacts of higher education institutions on sustainable development—An analysis of tools and indicators. *Sustainability*, *11* (1), 59. https://doi.org/10.3390/su11010059

Fischer, D., Jenssen, S., & Tappeser, V. (2015). Getting an empirical hold of the sustainable university: A comparative analysis of evaluation frameworks across 12 contemporary sustainability assessment tools. *Assessment & Evaluation in Higher Education*, *40* (6), 785–800. https://doi.org/10.1080/02602938.2015.1043234

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55* (1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Fry, D., Mooney, S. J., Rodríguez, D. A., Caiaffa, W. T., & Lovasi, G. S. (2020). Assessing Google Street View image availability in Latin American cities. *Journal of Urban Health*, *97*(4), 552–560. https://doi.org/10.1007/s11524-019-00408-7

Fuentes, J. E., Garcia, C. E., & Olaya, R. A. (2022). Estimation of the setting and infrastructure criterion of the UI GreenMetric ranking using unmanned aerial vehicles. *Sustainability*, *14* (1), 46. https://doi.org/10.3390/su14010046

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, *114* (50), 13108–13113. https://doi.org/10.1073/pnas.1700035114

Han, Y., Zhong, T., Yeh, A. G. O., Zhong, X., Chen, M., & Lü, G. (2023). Mapping seasonal changes of street greenery using multi-temporal street-view images. *Sustainable Cities and Society*, *92*, 104498. https://doi.org/10.1016/j.scs.2023.104498

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hou, Y., & Biljecki, F. (2022). A comprehensive framework for evaluating the quality of street view imagery. *International Journal of Applied Earth Observation and Geoinformation*, *115*, 103094. https://doi.org/10.1016/j.jag.2022.103094

İncekara, A. H., Yaprak Başaran, E., & Şeker, D. Z. (2022). The potential contribution of remotely sensed images for GreenMetric ranking. *International Journal of Environment and Geoinformatics*, *9* (4), 138–
150. https://doi.org/10.30897/ijegeo.1141366

Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8. https://github.com/ultralytics/ult ralytics

Jocher, G., & Qiu, J. (2024). Ultralytics YOLO11. https://github.com/ultralytics/ultralytics

Kamal, A. S. M., & Asmuss, M. (2013). Benchmarking tools for assessing and tracking sustainability in higher educational institutions: Identifying an effective tool for the University of Saskatchewan. *International Journal of Sustainability in Higher Education*, *14* (4), 449–465. https://doi.org/10.1108/IJSHE-08-2 011-0052

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, *4* (9), 100804. https://doi.org/10.1016/j.patter.2023.100804

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30* (pp. 3146–3154). Curran Associates.

Kim, J., & Jang, K. M. (2023). An examination of the spatial coverage and temporal variability of Google Street View (GSV) images in small- and medium-sized cities: A people-based approach. *Computers, Environment and Urban Systems*, *102*, 101956. https://doi.org/10.1016/j.compenvurbsys.2023.101956

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46* (1), 33–50. https://doi.org/10.2307/1913643

Krylov, V. A., Kenny, E., & Dahyot, R. (2018). Automatic discovery and geotagging of objects from street view imagery. *Remote Sensing*, *10* (5), 661. https://doi.org/10.3390/rs10050661

Lad, N., & Akerlof, K. L. (2022). Assessing campus sustainability literacy and culture: How are universities doing it and to what end? *Frontiers in Sustainability*, *3*, 927294. https://doi.org/10.3389/frsus.2022.927294

Lambrechts, W. (2015). The contribution of sustainability assessment to policy development in higher education. *Assessment & Evaluation in Higher Education*, *40* (6), 801–816. https://doi.org/10.1080/02602938.2015.1040719

Larkin, A., & Hystad, P. (2019). Evaluating street view exposure measures of visible green space for health research. *Journal of Exposure Science & Environmental Epidemiology*, *29* (4), 447–456. https://doi.or g/10.1038/s41370-018-0017-1

Lauder, A., Sari, R. F., Suwartha, N., & Tjahjono, G. (2015). Critical review of a global campus sustainability ranking: GreenMetric. *Journal of Cleaner Production*, *108*, 852–863. https://doi.org/10.1016/j.jclepro.2015.02.080

Laumer, D., Lang, N., van Doorn, N., Mac Aodha, O., Perona, P., & Wegner, J. D. (2020). Geocoding of trees from street addresses and street-level images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 125–136. https://doi.org/10.1016/j.isprsjprs.2020.02.001

Li, X. (2021). Examining the spatial distribution and temporal change of the green view index in New York City using Google Street View images and deep learning. *Environment and Planning B: Urban Analytics and City Science*, *48* (7), 2039–2054. https://doi.org/10.1177/2399808320962511

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, *14* (3), 675–685. https://doi.org/10.1016/j.ufug.2015.06.006

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014* (pp. 740–755). https://doi.org/10.1007/978-3-319-10602-1_48

Lozano, R. (2006). A tool for a Graphical Assessment of Sustainability in Universities (GASU). *Journal of Cleaner Production*, *14* (9–11), 963–972. https://doi.org/10.1016/j.jclepro.2005.11.041

Lozano, R., Ceulemans, K., Alonso-Almeida, M., Huisingh, D., Lozano, F. J., Waas, T., Lambrechts, W., Lukman, R., & Hugé, J. (2015). A review of commitment and implementation of sustainable development in higher education: Results from a worldwide survey. *Journal of Cleaner Production*, *108*, 1–18. https://doi.org/10.1016/j.jclepro.2014.09.048

Lukman, R., Krajnc, D., & Glavič, P. (2010). University ranking using research, educational and environmental indicators. *Journal of Cleaner Production*, *18* (7), 619–628. https://doi.org/10.1016/j.jclepro.2009.09.015

Minutolo, M. C., Ivanova, A., & Cong, M. (2021). Signaling sustainability: Impact that learning how to report has on enrollment, endowment and emissions of North American higher education institutions. *Sustainability Accounting, Management and Policy Journal*, *12* (5), 1140–1158. https://doi.org/10.1108/SAMPJ-06-2020-0224

Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics*, *41* (12), 1513–1523. https://doi.org/10.1080/00036840601019364

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore – Predicting the perceived safety of one million streetscapes. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 793–799). https://doi.org/10.1109/CVPRW.2014.121

Nassar, A. S., D'Aronco, S., Lefèvre, S., & Wegner, J. D. (2020). GeoGraph: Graph-based multi-view object detection with geometric cues end-to-end. In *Computer Vision – ECCV 2020* (pp. 488–504). https://doi.org/10.1007/978-3-030-58571-6_29

Neuhold, G., Ollmann, T., Rota Bulò, S., & Kontschieder, P. (2017). The Mapillary Vistas dataset for semantic understanding of street scenes. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 5000–5009). https://doi.org/10.1109/ICCV.2017.534

OECD and European Commission Joint Research Centre. (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. *OECD Publishing*. https://doi.org/10.1787/9789264043466-en

Orr, D. W. (1993). Architecture as pedagogy. *Conservation Biology*, *7*(2), 226–228. https://doi.org/10.1046/j.1523-1739.1993.07020226.x

Orr, D. W. (2021, February 21). Building as pedagogy: Oberlin's Adam Joseph Lewis Center. *Buildings & Cities*, Research Pathways. https://www.buildingsandcities.org/insights/research-pathways/ building-pedagogy.html

Palecki, M., Durre, I., Applequist, S., Arguez, A., & Lawrimore, J. (2021). U.S. Climate Normals 2020: U.S. Monthly Climate Normals (1991–2020). *NOAA National Centers for Environmental Information*. https://doi.org/10.25921/wck8-er13

Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176* (3), 609–634. https://doi.org/10.1111/ j.1467-985X.2012.01059.x

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31* (pp. 6638–6648). Curran Associates.

Qin, Y., Wu, X., Yu, T., & Jiang, S. (2025). Enhancing student-centered walking environments on university campuses through street view imagery and machine learning. *PLOS ONE*, *20* (4), e0321028. https:

//doi.org/10.1371/journal.pone.0321028

QS Quacquarelli Symonds. (2026). QS Sustainability Rankings. *QS Support Centre article*. https://support.qs.com/hc/en-gb/articles/8551503200668-QS-Sustainability-Rankings

Ranftl, R., Bochkovskiy, A., & Koltun, V. (2021). Vision Transformers for Dense Prediction. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 12159–12168). https://doi.org/10.1109/ICCV48922.2021.01196

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press. https://gaussianprocess.org/gpml/

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40* (8), 913–929. https://doi.org/10.1111/ecog.02881

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168* (2), 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation*, *12* (5), 1207–1245. https://doi.org/10.1162/089976600300015565

Seiferling, I., Naik, N., Ratti, C., & Proulx, R. (2017). Green streets – Quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, *165*, 93–101. https:

//doi.org/10.1016/j.landurbplan.2017.05.010

Shriberg, M. (2002). Institutional assessment tools for sustainability in higher education: Strengths, weak-nesses, and implications for practice and theory. *International Journal of Sustainability in Higher Edu-cation*, *3* (3), 254–270. https://doi.org/10.1108/14676370210434714

Stubbings, P., Peskett, J., Rowe, F., & Arribas-Bel, D. (2019). A hierarchical urban forest index using street-level imagery and deep learning. *Remote Sensing*, *11* (12), 1395. https://doi.org/10.3390/rs11121395

Times Higher Education. (2025). Impact Rankings 2025: Methodology. *Web page*. https://www.timeshig hereducation.com/world-university-rankings-news/impact-rankings-2025-methodology

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, *1*, 211–244. https://jmlr.org/papers/v1/tipping01a.html

U.S. Geological Survey. (2026). USGS EROS Archive – Aerial Photography – National Agriculture Imagery Program (NAIP). *Web page*. https://usgs.gov/centers/eros/science/usgs-eros-archive-aeria l-photography-national-agriculture-imagery-program-naip

Universitas Indonesia. (2026). UI GreenMetric Sustainable University Rankings: Guideline 2026. https: //uigreenmetric.com/wp-content/uploads/2026/06/2026_Guideline_UI-GreenMetric-SUR-eng-v 2.pdf

University of Louisville. (n.d.). Ratings & rankings. *Sustainability Council*. https://louisville.edu/sus tainability/sustainability-council/ratings-rankings

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, *7*, 91. https://doi.org/10.1186/1471-2105-7-91

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, *180*, 68–77. https://doi.org/10.1016/j.neuroimage.2017.06.061

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N., Jr., Fenimore, C., Gleason, K., & Arndt, D. (2014a). Improved historical temperature and precipitation time series for U.S. climate divisions. *Journal of Applied Meteorology and Climatology*, *53* (5), 1232–1251. https://doi.org/10.117 5/JAMC-D-13-0248.1

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N., Jr., Fenimore, C., Gleason, K., & Arndt, D. (2014b). NOAA Monthly U.S. Climate Gridded Dataset (NClimGrid), Version 1. *NOAA National Centers for Environmental Information*. https://doi.org/10.7289/V5SX6B56

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOv10: Real-time end-to-end object detection. In *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*. arXiv:2405.14458. https://arxiv.org/abs/2405.14458

Wegner, J. D., Branson, S., Hall, D., Schindler, K., & Perona, P. (2016). Cataloging public objects using aerial and street-level images – Urban trees. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6014–6023). https://doi.org/10.1109/CVPR.2016.647

Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, *22* (158), 209–212. https://doi.org/10.1080/01621459.1927.1050 2953

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/10.101 6/S0893-6080(05)80023-1

Wu, T., Lin, D., Chen, Y., & Wu, J. (2025). Integrating street view images, deep learning, and sDNA for evaluating university campus outdoor public spaces: A focus on restorative benefits and accessibility. *Land*, *14*(3), 610. https://doi.org/10.3390/land14030610

Xia, J., Yokoya, N., Adriano, B., & Broni-Bediako, C. (2023). OpenEarthMap: A Benchmark Dataset for Global High-Resolution Land Cover Mapping. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 6243–6253). https://doi.org/10.1109/WACV56688.2023.00619

Yin, D., & Hirata, T. (2025). Comprehensive comparative analysis and innovative exploration of green view index calculation methods. *Land*, *14*(2), 289. https://doi.org/10.3390/land14020289

Yuan, X., & Yu, L. (2026). Green campus assessment supported by remote sensing big data: A review of concepts, tools, evidence pathways, and future directions. *Sustainability*, *18* (14), 7437. https://doi.org/10.3390/su18147437

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*, *180*, 148–160. https://doi.org/10.1016/j.landurbplan.2018.08.020

Zhang, W., Witharana, C., Li, W., Zhang, C., Li, X., & Parent, J. (2018). Using deep learning to identify utility poles with crossarms and estimate their locations from Google Street View images. *Sensors*, *18* (8), 2484. https://doi.org/10.3390/s18082484

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84* (3), 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9

Zhao, T., Liang, X., Biljecki, F., Tu, W., Cao, J., Li, X., & Yi, S. (2025). Quantifying seasonal bias in street view imagery for urban form assessment: A global analysis of 40 cities. *Computers, Environment and Urban Systems*, *120*, 102302. https://doi.org/10.1016/j.compenvurbsys.2025.102302

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ADE20K dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5122–5130). https://doi.org/10.1109/CVPR.2017.544

Supplementary material

Note S1. Data dictionary

Shares are fractions (0–1) unless a name ends in _pct. Suffix _A = all panoramas in the window, equal weight; _B = overlap-corrected (one panorama per 20 m unit; repeated 27.7° strip removed for detections; line-of-bearing merging).

The file campus_windows_dataset has one row per campus, window rule, and window kind.

Column	Meaning
uniqueid	HIFLD/IPEDS campus identifier
rule	observation-window rule (bimonthly_anchored in the paper)
window_kind	primary (highest coverage), leaf_on (best window entirely in May–Sep), stars ($\geqslant$ 20% coverage, closest to the STARS submission)
window_id, months	first month of the window; its months (YYYY-MM;YYYY-MM)
reference_month, reference_year	unit-weighted median month and its year
n_panoramas_A, n_panoramas_B	valid panoramas in the window; panoramas kept in Version B
n_units	20 m sampling units observed in the window
coverage_A, coverage_B	share of coverable grid points within 20 m of a panorama in the window
share_hires_A/B	share of panoramas from 16,384-px canvases
veg_tree/grass/plant/total_A/B	mean vegetation pixel share of the cropped image
vegd_*_A/B	depth-weighted vegetation share
road/sidewalk/building_A/B	DeepLabV3+ shares: A as stored (share of the whole canvas, padding included); B share of the true image after removing the padding contribution (Section 4.3)
sidewalk_road_ratio_A/B	sidewalk / (sidewalk + road)
vehicles_50, persons_50, bicycles_50 (_A/B)	detections per panorama at confidence $\geqslant$ 0.50
<obj>_dets_per_pano_A, <obj>_presence_A	detections per panorama; share of panoramas with a detection (confidence $\geqslant$ 0.80)
<obj>_objects_per100u_B, <obj>_presence_units_B	merged objects per 100 units; share of units with a detection
<obj>_dedup_ratio	detections per merged object

<obj> is bikerack, bin, busstop or evcharger. Object measures failed screening and are descriptive only.

The file stars_detail_harmonized has one row per STARS campus.

Column	Meaning
report_url	dated STARS report page used
stars_version	2.2 or 3.0 (one 2.x report without a printed version coded 2.2)
overall_score_page	overall score on the page (stars_overall in models)
check, check_diff	overall score recomputed from subcategory points and its difference from the page (3.0: earned incl. innovation / available excl. innovation; 2.2: earned / available excl. innovation × 100 + innovation)
ops_pct	Operations category, % of available points
transport_pct	Transportation subcategory, %
grounds_pct	2.2 Grounds subcategory %; 3.0 Ecologically Managed Grounds credit %
academics_pct	Curriculum + Research, % (negative control)

The index outputs are vcei_goalposts.csv, with each indicator's dimension, direction, and 5th and 95th percentile goalposts, and vcei_scores.csv, with the geometric and arithmetic index, rank, 90% rank interval, and dimension scores for each campus, window, and version.

The panorama-level files can be shared without the images: pano_attributes (identifier, latitude, longitude, vehicle heading, pitch, roll, and capture year and month), pano_campus_links (campus, panorama, grid points within 20 m, and minimum distance), image_inventory_with_padding (canvas type, content fraction, and crop box), and pano_indicators (per-panorama vegetation shares and detection counts).

The overhead block is in naip_extended.parquet, with one row per campus and the 38 nx features of Table S8 averaged over every NAIP acquisition (step 14b; the state check is in naip_state_variance.csv, step 14c). Wide-search results by split and learner (step 19) are in search/<outcome>___campus_ext___<set>.csv and search/<outcome>___augmented_ext___<set>.csv, and the matching _params.csv files hold the setting chosen in every training fold. The feature sets are img (the features of the earlier search), imgx (with the full overhead block), imgxc (the overhead block without image metadata), and inst_imgx (with institutional covariates). The comparison on identical splits is in table9_wide_search.csv (step 19b).

Tables S1–S8

Table S1: Observation-window rules compared (252 campuses).

Rule	Windows per campus (mean)	Best-window coverage, A (median)	Best-window coverage, B (median)	Maximum span (months)	Best window crosses May/Sep boundary	Campus-years with a window ≥ 50%	Campus-years ≥ 25%
Two-month anchored (primary)	20.2	0.78	0.75	2	12.3%	1,024	1,853
Fixed two-month blocks	22.2	0.76	0.73	2	10.3%	926	1,811
Calendar month	28.2	0.75	0.71	1	0.0%	829	1,737
Calendar quarter	19.4	0.78	0.75	3	8.3%	996	1,850
Meteorological season	19.3	0.78	0.75	3	10.0%	1,016	1,844

For reference, pooling every capture from 2017–2023 reaches a median coverage of 0.80, and pooling all years of each calendar month reaches 0.94, but such composites mix up to 17 years of change and are not used as observation windows.

Table S2: Out-of-sample R^2 by learner family and pipeline.

		Best learner in the family								Pipelines fixed in advance				
Approach	n	Linear	Poly-nomial	Splines	Kernel	k-NN	Trees	Boost-ing	Neural net	Nested selection	Top-5 average	Stack-ing	All-learner median	All-learner average
STARS overall														
Institutional covariates	128	0.05	0.03	0.02	0.05	−0.08	0.02	−0.03	0.01	0.03	0.03	0.03	—	0.05
Imagery, primary window	128	0.03	0.04	−0.02	−0.02	−0.03	0.02	−0.04	0.00	−0.02	0.02	−0.01	—	0.02
Imagery, STARS-matched window	127	0.10	0.04	−0.03	0.10	0.00	0.07	0.02	0.11	0.01	0.07	0.01	—	0.08
Institutional covariates, STARS-matched sample	127	0.05	0.01	0.00	0.04	−0.09	0.01	−0.02	0.01	−0.02	0.02	0.00	—	0.04
Institutional + imagery, STARS-matched window	127	0.11	0.04	0.02	0.11	−0.01	0.06	0.01	0.10	0.08	0.07	0.06	—	0.08
Ten measures, Version A, primary window	128	0.00	0.00	−0.02	−0.01	−0.07	−0.07	−0.29	0.00	−0.06	−0.02	−0.03	—	−0.04
Ten measures, Version B, primary window	128	−0.01	0.00	−0.02	−0.01	−0.07	−0.03	−0.12	0.00	−0.10	−0.07	−0.05	—	−0.01
Imagery, multi-window average	128	0.10	0.00	0.02	0.11	0.00	0.04	−0.02	0.02	0.00	0.06	0.01	—	0.08
Imagery, window-augmented	128	0.11	—	—	0.11	0.08	0.07	0.04	—	0.09	0.10	0.07	—	0.10
Image content only, window-augmented	128	0.06	—	—	0.08	0.03	0.06	0.03	—	0.06	0.06	0.04	—	0.08
Institutional + imagery, window-augmented	128	0.11	—	—	0.11	0.08	0.08	0.04	—	0.10	0.10	0.08	—	0.11
Wide search: imagery, multi-window	128	0.13	0.03	−0.04	0.09	0.00	0.01	0.05	−0.40	−0.08	0.02	0.02	0.10	0.10
Wide search: imagery, window-augmented	128	0.16	0.07	0.06	0.13	0.09	0.10	0.09	0.06	0.09	0.12	0.11	0.12	0.12
Wide search: imagery + full overhead block, multi-window	128	0.10	0.04	−0.01	0.09	0.02	0.03	0.05	−0.12	0.01	0.04	0.01	0.07	0.08
Wide search: imagery + full overhead block, window-augmented	128	0.08	0.03	0.03	0.08	0.06	0.06	0.03	−0.16	0.05	0.07	0.04	0.08	0.08
Wide search: imagery + overhead block without metadata, multi-window	128	0.12	0.04	−0.01	0.08	0.04	0.05	0.03	−0.25	−0.03	0.04	−0.01	0.09	0.09
Wide search: institutional + imagery + full overhead, multi-window	128	0.09	0.04	−0.01	0.08	0.03	0.05	0.04	−0.26	0.02	0.05	0.05	0.08	0.08
Operations														
Institutional covariates	107	−0.01	−0.01	−0.06	−0.01	−0.15	−0.03	−0.08	−0.01	−0.07	−0.06	−0.05	—	
Imagery, primary window	107	−0.02	−0.02	−0.06	0.00	−0.05	−0.03	−0.10	−0.07	−0.11	−0.07	−0.08	−0.02 —	
Imagery, STARS-matched window	106	0.01	−0.02	−0.04	0.01	−0.03	0.01	−0.09	−0.02	−0.07	−0.03	−0.05	−0.02 —	−0.01
Institutional covariates, STARS-matched sample	106	−0.02	0.00	−0.04	−0.03	−0.12	−0.02	−0.10	0.00	−0.10	−0.05	−0.06	—	
Institutional + imagery, STARS-matched window	106	0.00	−0.02	−0.05	0.01	−0.08	0.00	−0.09	−0.04	−0.05	−0.03	−0.05	−0.02 —	−0.01
Ten measures, Version A, primary window	107	0.03	−0.02	−0.04	0.01	−0.01	−0.04	−0.23	0.00	−0.06	−0.03	−0.03	—	
Ten measures, Version B, primary window	107	0.00	−0.01	−0.02	0.01	−0.03	0.00	−0.15	0.00	−0.12	−0.06	−0.06	−0.02 —	
													0.00	

Imagery, multi-window average	107	0.09	0.05	0.01	0.07	0.04	0.07	−0.02	0.03	−0.02	0.05	0.03	—	
Imagery, window-augmented	107	0.08	—	—	0.06	0.04	0.05	0.01	—	0.05	0.05	0.02	0.07 —	0.07
Image content only, window-augmented	107	0.07	—	—	0.06	0.02	0.05	0.02	—	0.04	0.06	0.03	— 0.07	

Approach	n	Linear	Poly-nomial	Splines	Kernel	k-NN	Trees	Boost-ing	Neural net	Nested selection	Top-5 average	Stack-ing	All-learner median	All-learner average
Institutional + imagery, window-augmented	107	0.07	—	—	0.06	0.03	0.04	0.02	—	0.04	0.04	0.02	—	0.06
Wide search: imagery, multi-window	107	0.09	0.11	−0.02	0.09	0.00	0.08	0.06	−0.33	0.01	0.04	0.06	0.05	0.07
Wide search: imagery + full overhead block, multi-window	107	0.05	0.02	−0.03	0.06	−0.01	0.09	0.08	−0.52	−0.14	−0.03	−0.08	0.04	0.04
Wide search: imagery + full overhead block, window-augmented	107	0.05	0.00	0.02	0.03	−0.05	0.03	−0.07	−0.37	−0.17	−0.06	−0.18	0.01	−0.02
Wide search: imagery + overhead block without metadata, multi-window	107	0.07	0.04	−0.02	0.06	0.02	0.10	0.09	−0.41	−0.01	0.01	0.01	0.06	0.06
Wide search: institutional + imagery + full overhead, multi-window	107	0.07	0.01	−0.03	0.06	−0.01	0.08	0.06	−0.50	−0.07	−0.01	−0.03	0.04	0.03
Transportation														
Institutional covariates	107	−0.01	0.00	−0.06	0.01	−0.14	−0.04	−0.13	0.00	−0.02	−0.02	−0.02	—	
Imagery, primary window	107	0.01	0.00	0.01	0.01	−0.02	−0.05	−0.18	0.01	−0.02	−0.01	0.00	−0.02 —	
Imagery, STARS-matched window	106	0.02	0.02	−0.01	−0.03	−0.04	0.04	0.00	0.01	−0.05	−0.03	−0.03	−0.02 —	0.01
Institutional covariates, STARS-matched sample	106	−0.01	−0.02	−0.06	0.01	−0.12	−0.03	−0.11	−0.02	−0.03	−0.03	−0.03	— −0.02	
Institutional + imagery, STARS-matched window	106	0.02	0.01	0.00	−0.02	−0.05	0.02	−0.02	0.02	−0.10	−0.02	−0.04	—	0.01
Ten measures, Version A, primary window	107	0.05	0.03	0.02	0.02	0.00	−0.03	−0.22	0.01	−0.02	0.01	0.00	—	−0.01
Ten measures, Version B, primary window	107	0.05	0.02	0.03	0.03	0.02	−0.01	−0.12	0.03	−0.03	0.02	−0.01	—	0.02
Imagery, multi-window average	107	0.08	0.08	0.06	0.06	0.05	0.05	−0.06	0.09	0.05	0.05	0.02	—	0.06
Imagery, window-augmented	107	0.10	—	—	0.09	0.07	0.07	−0.02	—	0.08	0.09	0.05	—	0.08
Image content only, window-augmented	107	0.09	—	—	0.09	0.08	0.08	−0.03	—	0.08	0.09	0.07	—	0.08
Institutional + imagery, window-augmented	107	0.09	—	—	0.08	0.06	0.00	−0.08	—	0.06	0.08	0.04	—	0.04
Wide search: imagery, multi-window	107	0.09	0.10	0.03	0.08	0.07	0.07	0.03	−0.35	0.03	0.06	0.04	0.07	0.07
Wide search: imagery + full overhead block, multi-window	107	0.08	0.09	0.05	0.09	0.00	0.12	0.12	−0.36	0.03	0.11	0.09	0.09	0.09
Wide search: imagery + full overhead block, window-augmented	107	0.11	0.13	0.12	0.11	−0.01	0.05	0.00	−0.13	0.13	0.13	0.10	0.08	0.08
Wide search: imagery + overhead block without metadata, multi-window	107	0.09	0.09	0.04	0.07	0.01	0.12	0.12	−0.31	0.09	0.12	0.09	0.09	0.09
Wide search: institutional + imagery + full overhead, multi-window	107	0.08	0.09	0.06	0.08	−0.01	0.09	0.12	−0.67	0.08	0.11	0.06	0.09	0.08
Grounds														
Institutional covariates	107	0.05	0.05	0.05	0.00	0.03	−0.04	−0.09	0.00	−0.08	−0.05	−0.04	—	
Imagery, primary window	107	0.00	0.00	−0.06	−0.03	−0.11	−0.05	−0.08	−0.05	−0.05	−0.03	−0.02	0.00 —	
Imagery, STARS-matched window	106	0.05	−0.01	−0.09	−0.02	−0.05	−0.05	−0.11	−0.01	−0.03	−0.03	−0.02	−0.09 —	−0.01

Institutional covariates, STARS-matched sample	106	0.00	0.00	0.00	0.00	0.02	−0.06	−0.08	−0.01	−0.06	−0.03	−0.02	—	−0.01
Institutional + imagery, STARS-matched window	106	−0.03	−0.01	−0.10	−0.02	−0.06	−0.05	−0.08	−0.04	−0.08	−0.06	−0.03	—	−0.01

Approach	n	Linear	Poly-nomial	Splines	Kernel	k-NN	Trees	Boost-ing	Neural net	Nested selection	Top-5 average	Stack-ing	All-learner median	All-learner average
Ten measures, Version A, primary window	107	0.00	−0.05	−0.05	0.00	−0.01	−0.06	−0.15	−0.01	−0.08	−0.04	−0.05	—	−0.03
Ten measures, Version B, primary window	107	0.00	−0.02	−0.01	0.00	−0.04	−0.08	−0.21	−0.03	−0.06	−0.05	−0.06	—	−0.04
Imagery, multi-window average	107	−0.03	0.00	−0.04	−0.01	−0.01	−0.01	−0.18	−0.05	−0.05	−0.05	−0.05	—	−0.02
Imagery, window-augmented	107	−0.02	—	—	−0.02	−0.03	−0.01	−0.15	—	−0.03	−0.02	0.00	—	−0.03
Image content only, window-augmented	107	−0.01	—	—	−0.02	−0.02	−0.01	−0.18	—	−0.04	−0.01	0.00	—	−0.02
Institutional + imagery, window-augmented	107	−0.04	—	—	−0.02	−0.04	−0.10	−0.19	—	−0.06	−0.03	−0.02	—	−0.04
Wide search: imagery, multi-window	107	0.02	0.03	−0.03	0.01	0.00	0.02	0.00	−0.67	−0.07	−0.01	−0.01	0.00	0.02
Wide search: imagery + full overhead block, multi-window	107	0.00	0.02	−0.03	0.01	−0.03	0.01	−0.02	−0.41	−0.06	−0.02	−0.04	−0.01	0.00
Wide search: imagery + full overhead block, window-augmented	107	0.00	0.00	0.00	−0.01	−0.09	−0.06	−0.10	−0.23	−0.12	−0.04	−0.04	−0.02	−0.02
Wide search: imagery + overhead block without metadata, multi-window	107	0.00	0.03	−0.02	0.00	−0.03	0.00	−0.03	−0.34	−0.06	−0.02	−0.03	0.00	0.01
Wide search: institutional + imagery + full overhead, multi-window	107	0.00	0.03	−0.05	0.01	−0.02	0.01	−0.03	−0.36	−0.06	−0.01	−0.03	0.00	0.00
Curriculum + Research (control)														
Institutional covariates	107	0.03	0.01	0.01	0.01	−0.13	−0.05	−0.03	0.00	−0.02	0.00	0.03	—	0.02
Imagery, primary window	107	0.06	0.03	−0.01	0.00	−0.03	0.01	−0.06	0.01	−0.02	0.00	−0.02	—	0.02
Imagery, STARS-matched window	106	0.12	0.06	0.00	0.12	0.04	0.09	0.11	0.09	0.06	0.10	0.09	—	0.11
Institutional covariates, STARS-matched sample	106	0.03	0.00	0.00	0.00	−0.14	−0.05	−0.03	−0.02	−0.09	−0.01	0.01	—	0.01
Institutional + imagery, STARS-matched window	106	0.12	0.02	0.02	0.12	0.03	0.08	0.11	0.10	0.07	0.10	0.08	—	0.11
Ten measures, Version A, primary window	107	0.00	0.00	−0.02	0.00	−0.07	−0.07	−0.25	0.00	−0.21	−0.03	−0.12	—	−0.03
Ten measures, Version B, primary window	107	0.00	0.00	−0.03	−0.01	−0.06	−0.03	−0.09	0.00	−0.05	−0.02	−0.01	—	−0.01
Imagery, multi-window average	107	0.05	0.00	−0.04	0.05	−0.05	0.05	0.01	0.01	−0.04	0.00	0.00	—	0.05
Imagery, window-augmented	107	0.08	—	—	0.07	0.04	0.04	−0.09	—	0.03	0.07	0.02	—	0.05
Image content only, window-augmented	107	0.04	—	—	0.04	0.01	0.01	−0.10	—	0.02	0.03	−0.01	—	0.02
Institutional + imagery, window-augmented	107	0.08	—	—	0.06	0.06	0.02	−0.05	—	0.02	0.05	0.01	—	0.07
Wide search: imagery, multi-window	107	0.02	0.00	−0.03	0.05	−0.07	0.01	0.03	−0.64	−0.06	−0.05	−0.05	0.00	0.01
Wide search: imagery + full overhead block, multi-window	107	0.04	0.08	−0.03	0.01	−0.03	0.03	0.06	−0.18	−0.03	0.00	0.01	0.03	0.04
Wide search: imagery + full overhead block, window-augmented	107	0.04	0.07	−0.06	0.03	−0.02	0.06	0.04	−0.27	−0.06	0.01	−0.01	0.06	0.04
Wide search: imagery + overhead block without metadata, multi-window	107	0.05	0.11	−0.03	0.04	−0.02	0.05	0.05	−0.16	−0.02	0.04	0.04	0.05	0.06
Wide search: institutional + imagery + full overhead, multi-window	107	0.02	0.05	−0.03	0.01	−0.01	0.05	0.10	−0.25	−0.05	0.02	−0.02	0.03	0.04
QS Sustainability														
Institutional covariates	113	0.36	0.30	0.30	0.34	0.25	0.37	0.38	0.30	0.32	0.32	0.34	—	0.37

Approach	n	Linear	Poly-nomial	Splines	Kernel	k-NN	Trees	Boost-ing	Neural net	Nested selection	Top-5 average	Stack-ing	All-learner median	Al l-learner average
Imagery, primary window	113	0.24	0.23	0.21	0.19	0.17	0.20	0.17	0.22	0.23	0.23	0.20	—	0.24
Ten measures, Version A, primary window	113	0.08	−0.07	−0.04	0.06	0.02	0.00	−0.11	−0.02	0.00	0.04	0.03	—	0.02
Ten measures, Version B, primary window	113	0.04	−0.03	−0.04	0.07	−0.02	−0.02	−0.23	−0.04	−0.04	−0.01	0.00	—	−0.01
Imagery, multi-window average	113	0.19	0.21	0.11	0.16	0.08	0.17	0.16	0.15	0.14	0.15	0.16	—	0.20
Imagery, window-augmented	113	0.27	—	—	0.23	0.16	0.23	0.22	—	0.25	0.28	0.25	—	0.27
Image content only, window-augmented	113	0.19	—	—	0.18	0.13	0.18	0.15	—	0.18	0.20	0.19	—	0.20
Institutional + imagery, window-augmented	113	0.39	—	—	0.29	0.21	0.38	0.36	—	0.35	0.41	0.38	—	0.39
Wide search: imagery, multi-window	113	0.30	0.24	0.14	0.29	0.07	0.26	0.28	0.06	0.17	0.22	0.21	0.28	0.28
Wide search: imagery + full overhead block, multi-window	113	0.31	0.29	0.20	0.30	0.14	0.28	0.30	0.00	0.29	0.30	0.28	0.30	0.31
Wide search: imagery + full overhead block, window-augmented	113	0.32	0.27	0.30	0.31	0.26	0.23	0.22	0.07	0.26	0.29	0.29	0.32	0.33
Wide search: imagery + overhead block without metadata, multi-window	113	0.33	0.31	0.20	0.30	0.12	0.27	0.32	0.01	0.24	0.26	0.24	0.30	0.31
Wide search: institutional + imagery + full overhead, multi-window	113	0.37	0.41	0.32	0.37	0.21	0.41	0.39	0.17	0.28	0.36	0.35	0.39	0.40
Wide search: institutional + imagery + full overhead, augmented	113	0.39	0.40	0.39	0.39	0.29	0.38	0.36	0.19	0.29	0.35	0.33	0.40	0.41

Values are means of three repeats of 5-fold cross-validation, except in the wide-search rows, which use one random 5-fold split (the first of the three). A family's value is the highest R^2 among its learners, which overstates what choosing that learner would deliver; the pipelines (nested selection, the average of the five learners best in inner cross-validation, non-negative stacking, and the average of all learners) involve no choice on the test folds. Window-augmented models used nine learners, with no polynomial, spline, or neural-network learner. The wide search (Section 4.5) used 33 learners per campus and 31 window-augmented, each tuned by random search in every training fold; the median of all learners was computed for the wide search only, and the least-angle lasso could not be fitted where features outnumbered inner training campuses (R^2 0 there). Learner settings are in Tables S12–S14, and results for every learner, scheme, and window are in tableS_model_search_all.csv in the replication package.

Table S3: In-sample least-squares fit.

Model	n	Predictors	R^2	Adjusted R^2	p (F)	p (partial F)
STARSoverall						
Tenmeasures,VersionB	128	10	0.06	−0.02	0.624	
Tenmeasures,VersionA	128	10	0.05	−0.03	0.756	
Allimageryfeatures(multi-window)	128	42	0.39	0.08	0.168	
Institutionalcovariates	128	12	0.13	0.05	0.119	
Institutional+tenmeasures(B)	128	21	0.21	0.05	0.163	
					0.370	
Operations						
Tenmeasures,VersionB	107	10	0.10	0.00	0.437	
Tenmeasures,VersionA	107	10	0.09	0.00	0.479	
Allimageryfeatures(multi-window)	107	42	0.48	0.13	0.121	
Institutionalcovariates	107	12	0.15	0.05	0.154	
Institutional+tenmeasures(B)	107	21	0.27	0.10	0.089	0.151
Transportation						
Tenmeasures,VersionB	107	10	0.14	0.05	0.114	
Tenmeasures,VersionA	107	10	0.14	0.05	0.141	
Allimageryfeatures(multi-window)	107	42	0.40	0.00	0.478	
Institutionalcovariates	107	12	0.17	0.07	0.076	
Institutional+tenmeasures(B)	107	21	0.25	0.07	0.151	
					0.460	
Grounds						
Tenmeasures,VersionB	107	10	0.09	0.00	0.443	
Tenmeasures,VersionA	107	10	0.10	0.01	0.367	
Allimageryfeatures(multi-window)	107	42	0.43	0.06	0.286	
Institutionalcovariates	107	12	0.14	0.04	0.186	
Institutional+tenmeasures(B)	107	21	0.21	0.01	0.419	0.715
Curriculum+Research(control)						
Tenmeasures,VersionB	107	10	0.13	0.04	0.199	
Tenmeasures,VersionA	107	10	0.12	0.03	0.256	
Allimageryfeatures(multi-window)	107	42	0.42	0.03	0.376	
Institutionalcovariates	107	12	0.14	0.04	0.190	
Institutional+tenmeasures(B)	107	21	0.29	0.11	0.063	
Tenmeasures,VersionB	113	10	0.16	0.08	0.082	

QS Sustainability

R^2, adjusted R^2, and the p-value of the overall F-test; for institutional covariates plus the ten measures, the partial F-test of the ten measures. The ten measures are green view; road, sidewalk, and building shares; vehicles and persons per panorama; and bike racks, garbage bins, bus stops, and EV chargers, all in the primary window. Missing values were replaced by the median. No predictor is derived from a sustainability rating.

Table S3b. Standardized least-squares coefficients of the ten measures.

Measure	STARS overall	Opera-tions	Transpor-tation	Grounds	Curric. + Research	QS Sus-tainability
Greenery (green view)	−0.79 (0.69)	1.89 (0.46)	−3.05 (0.47)	−2.63 (0.61)	−3.02 (0.27)	4.66 (0.37)
Road share	−0.80 (0.61)	0.40 (0.84)	−3.85 (0.28)	0.00 (1.00)	−1.50 (0.47)	2.35 (0.66)
Sidewalk share	0.08 (0.95)	1.12 (0.50)	2.47 (0.41)	−5.46 (0.11)	0.28 (0.90)	−2.85 (0.51)
Building share	−0.46 (0.77)	0.00 (1.00)	1.11 (0.80)	−1.46 (0.72)	−1.08 (0.71)	9.49 (0.10)
Vehicles per panorama	−2.32 (0.24)	−1.09 (0.66)	−2.36 (0.53)	−5.87 (0.17)	−6.23 (0.05)	−5.08 (0.41)
Persons per panorama	0.60 (0.60)	1.93 (0.14)	3.13 (0.33)	7.16 (0.01)	−1.61 (0.25)	10.45 (0.32)
Bike racks	0.71 (0.54)	−0.50 (0.74)	−1.09 (0.70)	−4.05 (0.19)	1.13 (0.53)	0.45 (0.92)
Garbage bins	2.76 (0.21)	2.40 (0.31)	2.05 (0.64)	−0.71 (0.89)	7.30 (0.03)	−4.31 (0.48)
Bus stops	−2.10 (0.21)	−2.44 (0.20)	−1.05 (0.77)	4.78 (0.19)	−4.52 (0.03)	−1.36 (0.73)
EV chargers	1.86 (0.12)	1.59 (0.20)	0.56 (0.81)	−0.06 (0.98)	0.02 (0.99)	−1.68 (0.59)

Score points per standard deviation of each Version B measure, with HC3 p-values in parentheses. Score scales: STARS overall, 0–100 points; STARS sub-scores, percent of available points; QS Sustainability, 0–100.

Table S4: Goalposts of the Visible Campus Environment Index.

Dimension	**Indicator**	**Direction**	**Lower goalpost (5th pct)**	**Upper goalpost (95th pct)**	**Cam-puses**
Greenery	Street green view (share)	higher is better	0.035	0.188	249
Greenery	Street tree share	higher is better	0.018	0.143	249
Greenery	NAIP tree canopy (%)	higher is better	7.5	47.3	248
Greenery	NAIP vegetation (%)	higher is better	18.6	75.5	248
Pedestrian space	Street sidewalk share	higher is better	0.009	0.031	249
Pedestrian space	Sidewalk / (sidewalk + road)	higher is better	0.021	0.074	249
Car dominance	Vehicles per panorama	lower is better	0.6	5.6	249
Car dominance	Street road share (padding-corrected)	lower is better	0.374	0.429	249
Car dominance	NAIP roads + other paved (%)	lower is better	16.3	49.9	248

Version B, primary window. Campus-level index values with 90% rank intervals are in tableS_vcei_ranks.csv.

Table S5: Panorama-level measures by canvas type.

Measure	4,096 full	4,096 padded	16,384 full	16,384 padded
Green view	7.8 / 6.3 / 16.4	8.7 / 7.3 / 17.5	7.8 / 6.2 / 16.2	9.6 / 8.2 / 18.8
Road, stored	41.8 / 41.7 / 45.7	49.8 / 49.9 / 53.1	41.7 / 41.7 / 45.6	48.8 / 48.9 / 52.2
Road, corrected	41.8 / 41.7 / 45.7	41.4 / 41.6 / 46.0	41.7 / 41.7 / 45.6	41.3 / 41.5 / 46.4
Sidewalk, stored	1.9 / 1.5 / 4.1	1.5 / 1.1 / 3.1	1.9 / 1.5 / 4.1	1.4 / 1.1 / 3.0
Sidewalk, corrected	1.9 / 1.5 / 4.1	2.0 / 1.5 / 4.3	1.9 / 1.5 / 4.1	2.0 / 1.5 / 4.5
Building, stored	7.1 / 5.2 / 16.4	6.1 / 5.4 / 10.9	7.0 / 5.0 / 16.2	12.9 / 11.8 / 17.3
Building, corrected	7.1 / 5.2 / 16.4	6.8 / 5.7 / 13.5	7.0 / 5.0 / 16.2	7.0 / 5.4 / 13.6
Vehicles (count)	2.79 / 2 / 7 (28% zero)	2.68 / 2 / 7 (31% zero)	2.62 / 2 / 7 (30% zero)	2.41 / 1 / 6 (32% zero)
Persons (count)	0.14 / 0 / 0 (92% zero)	0.12 / 0 / 0 (92% zero)	0.13 / 0 / 0 (92% zero)	0.12 / 0 / 0 (92% zero)
Bike-rack detections	0.00 / 0 / 0 (100% zero)	0.00 / 0 / 0 (100% zero)	1.40 / 1 / 4 (40% zero)	1.23 / 1 / 3 (38% zero)
Bin detections	0.05 / 0 / 0 (95% zero)	0.04 / 0 / 0 (96% zero)	0.46 / 0 / 1 (67% zero)	0.40 / 0 / 1 (70% zero)

All valid panoramas. Shares are in percent and counts are per panorama, each given as mean / median / 90th percentile, followed by the share of panoramas with zero. Panoramas: 646,067 (4,096 full), 1,232,541 (4,096 padded), 209,483 (16,384 full), and 331,496 (16,384 padded). Vegetation is computed on the cropped true image in both versions.

Table S6: Median campus-level measures by Carnegie class, locale, and census region.

Group	n	Green view (%)	Road (%)	Sidewalk (%)	Building (%)	Vehicles / panorama	Persons / panorama	NAIP canopy (%)	NAIP paved (%)
Carnegie: Bacc. A&S	102	13.5	39.7	1.9	6.5	2.0	0.03	26.7	30.1
Carnegie: Other	18	7.7	41.6	1.6	5.9	2.1	0.03	13.2	38.5
Carnegie: R1	96	9.2	40.6	2.0	7.4	2.8	0.13	19.1	35.8
Carnegie: R2	33	8.2	40.8	2.1	6.8	2.8	0.09	16.8	42.7
Locale: Missing	11	8.7	41.8	1.6	5.1	1.6	0.03	17.0	35.8
Locale: Suburban	60	11.8	40.3	1.8	6.3	2.2	0.04	27.8	30.6
Locale: Town/rural	38	14.5	39.7	1.7	6.6	1.5	0.03	24.6	27.0
Locale: Urban	140	9.1	40.3	2.1	7.3	2.8	0.10	19.0	36.9
Region: Midwest	50	10.8	40.2	2.2	6.8	2.2	0.06	18.3	35.2
Region: Northeast	86	12.8	39.8	1.9	7.4	2.4	0.08	23.0	33.0
Region: South	66	11.6	40.5	2.0	5.9	2.4	0.06	27.2	33.1
Region: West	47	8.8	41.1	1.7	6.6	2.9	0.05	18.4	37.9

Version B, primary window.

Table S7: Outcome distributions and correlations among outcomes.

Outcome	STARS version	n	Mean(SD)	Min	25th	Median	75th	Max
STARSoverall	2.2	91	66.9(10.4)	37.1	62.3	67.3	72.5	86.9
STARSoverall	3.0	37	62.4(11.2)	40.5	54.2	65.6	68.2	87.0
STARSoverall	all	128	65.6(10.8)	37.1	59.7	66.7	71.9	87.0
Operations	2.2	70	51.1(11.1)	18.6	46.8	51.8	59.4	70.5
Operations	3.0	37	45.7(10.9)	23.4	38.8	45.0	50.2	73.8
Operations	all	107	49.3(11.3)	18.6	43.8	49.8	57.5	73.8
Transportation	2.2	70	55.8(20.5)	10.0	46.9	61.0	70.4	90.3
Transportation	3.0	37	39.5(20.0)	0.0	24.8	41.7	53.1	73.7
Transportation	all	107	50.2(21.7)	0.0	37.0	54.0	68.2	90.3
Grounds	2.2	70	67.6(26.1)	0.0	60.2	75.0	83.1	100.0
Grounds	3.0	37	53.3(22.6)	10.2	42.4	53.8	69.2	100.0
Grounds	all	107	62.6(25.8)	0.0	44.7	72.4	81.1	100.0
Curriculum+Research	2.2	70	76.2(13.4)	43.6	68.5	75.4	84.6	100.0
Curriculum+Research	3.0	37	69.0(14.2)	39.1	58.6	69.4	78.5	94.4
Curriculum+Research	all	107	73.7(14.0)	39.1	65.9	74.3	82.5	100.0
QSSustainability	all	113	37.1(31.5)	1.0	8.4	29.3	62.5	100.0

	STARS overall	Opera-tions	Transporta-tion	Grounds	Curriculum + Research	QS Sustain-ability
STARS overall	1.00	0.82	0.64	0.43	0.81	0.24
Operations	0.82	1.00	0.62	0.53	0.48	0.25
Transportation	0.64	0.62	1.00	0.33	0.42	0.27
Grounds	0.43	0.53	0.33	1.00	0.36	0.07
Curriculum + Research	0.81	0.48	0.42	0.36	1.00	0.14
QS Sustainability	0.24	0.25	0.27	0.07	0.14	1.00

Campuses with a usable primary window. Correlations are Spearman ρ.

Table S8: The 38 overhead land-cover features.

Feature	Mean (SD)	10th pct	Median	90th pct	ρ, street green view	ρ, STARS overall	ρ, QS	Variance explained by state (chance)
Grass (%)	22.2 (8.76)	11.5	22.7	33.6	0.32	−0.10	−0.13	0.15 (0.15)
Pavement (%)	23.6 (8.50)	13.4	23.3	33.6	−0.60	−0.13	−0.06	0.15 (0.15)
Road (%)	10.2 (3.05)	6.28	10.2	14.3	−0.44	−0.18	−0.43	0.18 (0.16)
Tree (%)	23.7 (12.3)	9.58	21.8	41.3	0.67	0.13	0.06	0.19 (0.15)
Water (%)	0.57 (0.98)	0.00	0.14	1.83	0.07	0.07	0.09	0.18 (0.15)
Cropland (%)	3.47 (4.80)	0.01	1.83	8.75	0.38	−0.01	0.16	0.23 (0.15)
Building (%)	16.1 (8.85)	6.86	14.0	27.6	−0.58	0.03	−0.04	0.15 (0.15)
Bare land (%)	0.14 (0.56)	0.00	0.00	0.20	−0.01	0.02	0.14	0.15 (0.16)
Vegetation (%)	49.4 (17.1)	26.9	50.0	71.0	0.65	0.07	0.06	0.15 (0.15)
Impervious (%)	49.9 (17.3)	28.4	49.4	72.6	−0.64	−0.08	−0.07	0.15 (0.15)
Vegetation to impervious (log 1+ratio)	0.76 (0.38)	0.32	0.71	1.26	0.64	0.08	0.07	0.14 (0.15)
Tree to impervious (log 1+ratio)	0.45 (0.31)	0.14	0.37	0.83	0.70	0.12	0.09	0.15 (0.15)
Building to vegetation (log 1+ratio)	0.36 (0.38)	0.09	0.25	0.67	−0.61	−0.01	−0.04	0.11 (0.15)
Roads, share of impervious cover (%)	21.8 (5.68)	14.5	22.4	29.0	0.34	−0.12	−0.24	0.19 (0.15)
Shannon diversity of land cover	1.56 (0.14)	1.39	1.58	1.69	0.07	−0.01	0.01	0.16 (0.15)
Shannon evenness	0.75 (0.07)	0.67	0.76	0.81	0.07	−0.01	0.01	0.16 (0.15)
Tree, largest patch (% of campus)	23.0 (17.6)	6.25	16.7	49.9	0.42	0.06	0.08	0.18 (0.15)
Grass, largest patch (% of campus)	15.8 (10.1)	6.42	13.1	29.0	−0.13	−0.04	0.06	0.11 (0.15)
Building, largest patch (% of campus)	12.4 (8.69)	5.11	10.3	22.3	0.08	−0.27	−0.19	0.23 (0.15)
Pavement, largest patch (% of campus)	18.9 (11.6)	8.92	16.1	29.8	−0.12	−0.04	−0.07	0.13 (0.15)
Vegetation, largest patch (% of campus)	35.6 (21.1)	14.0	30.4	68.2	0.45	−0.05	0.09	0.11 (0.15)
Impervious, largest patch (% of campus)	77.9 (18.9)	48.3	83.5	96.9	−0.34	−0.05	−0.21	0.17 (0.15)
Tree, patches per 100 ha	720.5 (544.5)	166.8	693.9	1,240.4	−0.10	−0.20	−0.38	0.31 (0.15)
Grass, patches per 100 ha	984.2 (599.5)	226.5	1,028.7	1,680.4	0.16	−0.18	−0.38	0.24 (0.15)
Building, patches per 100 ha	217.2 (254.3)	67.1	182.8	344.8	−0.13	−0.15	−0.29	0.20 (0.14)
Pavement, patches per 100 ha	910.8 (499.0)	212.6	984.2	1,480.9	0.02	−0.18	−0.35	0.21 (0.15)
Road, patches per 100 ha	172.3 (89.2)	55.4	168.6	287.2	0.10	−0.23	−0.32	0.21 (0.15)
Vegetation, patches per 100 ha	485.8 (481.4)	117.2	425.5	912.8	−0.29	−0.19	−0.33	0.23 (0.14)
Impervious, patches per 100 ha	215.2 (148.4)	51.0	195.7	369.3	0.35	−0.25	−0.32	0.20 (0.15)
Water, patches per 100 ha	6.58 (6.09)	0.98	4.76	15.0	0.01	−0.11	−0.21	0.19 (0.15)
Campus area (log ha)	4.38 (1.18)	3.09	4.34	5.96	0.14	0.27	0.30	0.19 (0.15)
Mean classification confidence	0.85 (0.03)	0.81	0.85	0.88	0.26	−0.19	−0.36	0.17 (0.15)
Mean normalized entropy	0.18 (0.03)	0.15	0.18	0.22	−0.22	0.20	0.35	0.19 (0.15)
Pixels with confidence < 0.6 (%)	14.1 (4.14)	10.1	13.4	19.5	−0.23	0.21	0.42	0.16 (0.15)
Mean brightness, red band	120.6 (15.6)	100.6	120.4	140.9	−0.43	−0.07	−0.08	0.37 (0.15)
Mean brightness, green band	128.2 (14.9)	107.6	128.4	146.0	−0.32	−0.09	−0.07	0.47 (0.15)
Mean brightness, blue band	115.3 (16.1)	93.8	115.4	136.1	−0.45	−0.02	−0.04	0.32 (0.15)
Number of acquisitions	3.16 (0.85)	2.00	3.00	4.00	−0.23	0.06	0.06	0.41 (0.15)

Features from the NAIP land-cover segmentation, averaged over every acquisition of a campus (2018–2023), for the 251 analysis campuses with NAIP imagery. ρ is the Spearman correlation across campuses (STARS overall: 128 campuses; QS: 113) and is descriptive only. Variance explained by state is η^2 across the 39 states (the District of Columbia counted as one), with its mean when state labels are permuted at random (200 permutations), the value expected by chance, in parentheses. NAIP is flown state by state, so brightness and the number of acquisitions differ by state. Shares of the eight classes sum to about 100%; vegetation is tree + grass + cropland, and impervious is building + road + pavement. NDVI cannot be computed because the tiles have no near-infrared band.

Note S2. Image-processing models and detector training

Table S9 lists every model that produced a measurement used in the paper, with the weights and settings it was run with. The vegetation, depth, road-sidewalk-building, person-vehicle, and overhead land-cover models were used with their published weights. Only the four special-object detectors were trained for this project.

Table S9: Image-processing models and settings.

Measure	Model	Weights (training data)	Settings
Vegetation (tree, grass, plant)	Mask2Former, Swin-L backbone	facebook/mask2former-swin-large-ade-semantic (ADE20K)	Black padding cropped; image resized to 4,000 × 2,000 px (Lanczos); ADE20K classes tree, grass, and plant; pixels kept within HSV ranges [20, 40, 30]–[95, 255, 255] for tree and grass and [20, 35, 30]–[92, 255, 255] for plant (OpenCV scale)
Depth-weighted vegetation	DPT-Large	Intel/dpt-large (MIX 6)	Relative depth predicted per pixel, min–max normalized within each image, and used as pixel weights for the vegetation share
Road, sidewalk, building	DeepLabV3+, MobileNetV2 backbone, output stride 16	Cityscapes (19 classes)	Full canvas, including any padding, downscaled to a longest side of 2,048 px; padding correction in Section 4.3
Persons, bicycles, vehicles	YOLO26-m	yolo26m.pt (COCO)	COCO classes 0–8 stored at confidence ≥ 0.25; the analysis uses ≥ 0.50, with vehicles = car, truck, bus, and motorcycle; default input size
Bike racks	YOLOv10-s, fine-tuned	Roboflow (Table S10)	Sliced inference (SAHI) on 2,048 × 2,048 px tiles with 20% overlap; model input 960 px; confidence ≥ 0.80; NMS IoU 0.70; at most 100 detections per image; tile predictions merged at IoU ≥ 0.50
Garbage bins	YOLOv8-s, fine-tuned	Roboflow (Table S10)	As for bike racks, with model input 768 px
Bus stops	YOLOv8-s, fine-tuned	Roboflow (Table S10)	As for bike racks, with model input 896 px
EV chargers	YOLO11-s segmentation, fine-tuned	Roboflow (Table S10)	As for bike racks, with model input 960 px
Overhead land cover	Mask2Former	mfaytin/mask2former-satellite (OpenEarthMap, 8 classes)	NAIP RGB tiles segmented in 512 px windows with 64 px overlap (pipeline version 2.0.1)

Confidence thresholds are those applied when the measures were computed. Person and vehicle detections were stored at a low threshold so that the analysis threshold could be chosen later.

The four detectors were fine-tuned from COCO-pretrained Ultralytics weights on Roboflow, using the training, validation, and test splits of the export (Table S10). For each object, we screened small (s) detection models from five YOLO generations (YOLOv8, YOLOv10, YOLO11, YOLO12, and YOLO26) and, where the labels were polygons, small segmentation models from three (YOLOv8, YOLO11, and YOLO26). We then retrained one to three of the screened architectures at medium (m) scale with more epochs (220–240) and more patience (40–45). In all, we made 33 training runs, one of which failed. Every run used the optimizer chosen automatically by Ultralytics, mixed precision, seed 42, fixed batch sizes (4 for small detection models, 2 for small segmentation and

medium detection models, and 1 for medium segmentation models), early stopping, and the same moderate augmentation: rotation up to 5°, translation 0.10, scaling 0.40, shear 2°, perspective 0.0005, horizontal flips with probability 0.5, hue, saturation, and value jitter of 0.015, 0.60, and 0.35, and mosaic with probability 0.80, switched off for the last 15 epochs.

We selected models on the validation set with a weighted score that favored image-level presence over box accuracy: 0.35 × presence F1 + 0.20 × presence recall + 0.15 × presence precision + 0.10 × presence accuracy + 0.05 × balanced accuracy + 0.15 × box mAP50–95. Presence means at least one detection above a threshold, which was tuned on the validation set (0.02 to 0.90 in steps of 0.01) to maximize F1. For every object, a small model from the screening won; no medium model did better.

Table S10: Detector training data and candidate models.

Object	Annotation	Training images (without object)	Validation images (without object)	Test images (without object)	Input size (px) / patience	Epochs (max)	Candidates screened (scale)	Refined (mscale)	Selected (validation score)
Bike racks	boxes and polygons	522 (6)	25 (0)	50 (0)	960	200/35	5	detection	YOLOv10-m	YOLOv10-s (0.90)
Garbage bins	boxes and polygons	543 (0)	26 (0)	52 (0)	768	180/30	5	detection	YOLOv8-m	YOLOv8-s (0.94)
Busstops	polygons	1,020 (263)	49 (20)	97 (25)	896	190/35	5	detection + segmentation	YOLOv8-m; YOLO26-m	YOLOv8-s (0.77)
EV chargers	polygons	819 (555)	43 (33)	111 (75)	960	200/35	5	detection + segmentation	YOLOv8-m (failed); YOLOv10-m; YOLO11-m segmentation	YOLO11-s segmentation (0.60)

Image counts in parentheses are images without the object. Validation score: the weighted selection score described above.

The selected detectors reached a box mAP50 of 0.38 for bike racks, 0.71 for garbage bins, 0.48 for bus stops, and 0.19 for EV chargers on their test sets (Table S11). The presence metrics look better than they are. The bike-rack and garbage-bin validation and test sets contain no image without the object, so a presence threshold of 0.02 yields perfect presence scores while saying nothing about false positives. Only the bus-stop and EV-charger sets contain such images, and there the specificity was
0.60 and 0.93. More important, these are curated images, mostly close-ups, and not tiles of street-level panoramas. On campus panoramas at a confidence of 0.80, a random screen of 24 detections per object found 4 of 96 correct (Section 5.4). We therefore treat the object measures as descriptive only.

Table S11: Detector accuracy on validation, test, and campus images.

Object	Evaluationset	Boxprecision	Boxrecall	mAP50	mAP50–95	Presencethreshold	Presenceaccuracy	Presenceprecision	Presencerecall	PresenceF1	Specificity
Bikeracks	Validation	0.79	0.54	0.55	0.30	0.02	1.00	1.00	1.00	1.00	—
Bikeracks	Test	0.65	0.34	0.38	0.19	0.02	1.00	1.00	1.00	1.00	—
Bikeracks	Campuspanoramas (screen)	0.00(0/24)	—	—	—	0.80	—	—	—	—	—
Garbagebins	Validation	0.92	0.78	0.85	0.57	0.02	1.00	1.00	1.00	1.00	—
Garbagebins	Test	0.75	0.73	0.71	0.37	0.02	1.00	1.00	1.00	1.00	—
Garbagebins	Campuspanoramas (screen)	0.08(2/24)	—	—	—	0.80	—	—	—	—	—
Busstops	Validation	0.49	0.38	0.35	0.16	0.22	0.84	0.78	1.00	0.88	0.60
Busstops	Test	0.63	0.52	0.48	0.23	0.22	0.84	0.87	0.92	0.89	0.60
Busstops	Campuspanoramas (screen)	0.08(2/24)	—	—	—	0.80	—	—	—	—	—
EVchargers	Validation	0.53	0.39	0.37	0.23	0.38	0.86	0.83	0.50	0.62	0.97
EVchargers	Test	0.41	0.20	0.19	0.09	0.38	0.76	0.74	0.39	0.51	0.93
EVchargers	Campuspanoramas (screen)	0.00(0/24)	—	—	—	0.80	—	—	—	—	—

Box metrics as reported by Ultralytics (precision and recall at the confidence that maximizes F1; mAP over IoU 0.50 and 0.50–0.95). Presence: an image counts as positive if it has a detection above the presence threshold, tuned on the validation set. Specificity: the share of images without the object that have no detection; —: no such images in the set. EV-charger masks: mAP50 0.33 and mAP50–95 0.23 (validation), 0.18 and 0.12 (test). Campus panoramas: the share of 24 randomly drawn detections at confidence ≥ 0.80 judged correct in the screen (Wilson 95% interval for all 96: 1.6–10.2%).

Note S3. Prediction models and their settings

All prediction models were fitted with scikit-learn 1.9.1, XGBoost 3.4.1, LightGBM 4.7.0, and CatBoost 1.2.10. Every learner used the same preprocessing, fitted inside each training fold. Numerical features were median-imputed, constant features were dropped, and the rest were Yeo–Johnson transformed and standardized; categorical covariates (Carnegie class, locale, and census region) were one-hot encoded. Forests and boosting received median-imputed features and one-hot covariates without further transformation. Learners marked with an asterisk in Tables S12–S14 were fitted to the outcome standardized within the training fold, and their predictions were transformed back. Stochastic learners used seed 0.

Outer evaluation used three repeats of 5-fold cross-validation over campuses (shuffled with seeds 1000, 1001, and 1002) and 5-fold cross-validation grouped by state. Within each outer training set, each learner was tuned by 5-fold cross-validation, the setting with the lowest inner mean squared error was refitted on the whole training set, and the inner out-of-fold predictions of that setting were kept for the pipelines. The pipelines were nested selection (the learner with the lowest inner error), the average of the five learners with the lowest inner error, stacking with non-negative weights fitted to the inner out-of-fold predictions, and the average of all learners. Table S12 lists the 27 learners of the earlier search with the grids searched; forests and boosting used fixed settings suited to about a hundred campuses.

Table S12: Learners of the earlier search and their settings.

Family	Learner	Fixed settings	Tuned in inner cross-validation
Linear	Ordinary least squares	Up to 12 curated features: window means of green view, depth-weighted green view, road, sidewalk, and building shares, and vehicles and persons per panorama (with log enrollment when covariates are included)	—
Linear	Ridge	α chosen by efficient leave-one-out cross-validation among 40 values from 0.01 to 10,000	—
Linear	Lasso*	—	$\alpha \in \{0.02, 0.05, 0.1, 0.2, 0.4\}$
Linear	Elastic net*	—	$\alpha \in \{0.05, 0.15, 0.4, 1.0\}$; L1 ratio $\in \{0.2, 0.5\}$
Linear	Bayesian ridge	Default priors	—
Linear	Huber regression*	—	$\alpha \in \{1, 10, 100\}$
Linear	Partial least squares	Unscaled	Components $\in \{1, 2, 3, 5\}$
Linear	Principal-component regression	Ridge on the components, α by leave-one-out among 30 values from 0.01 to 10,000	Components $\in \{3, 5, 10, 20\}$
Linear	Ridge after screening	Ridge α as for principal-component regression	k $\in \{5, 10, 20, 40\}$
Polynomial	Degree-2 ridge	Screening, degree-2 expansion, standardization; ridge α among 30 values from 0.1 to 100,000	k $\in \{5, 10, 20\}$
Polynomial	Degree-2 lasso*	Screening, degree-2 expansion, standardization	k $\in \{5, 10\}$; $\alpha \in \{0.05, 0.15, 0.4\}$
Polynomial	Degree-3 ridge	As degree-2 ridge, with degree 3	k $\in \{3, 5, 8\}$
Additive (splines)	Spline ridge	Cubic B-splines with 4 knots per screened feature; ridge α among 30 values from 0.1 to 100,000	k $\in \{5, 10, 20, 40\}$
Kernel	Kernel ridge, RBF*	—	$\alpha \in \{0.3, 1, 3\}$; $\gamma \in \{0.1, 0.3, 1\}$/p

Family	Learner	Fixed settings	Tuned in inner cross-validation
Kernel	Kernel ridge, polynomial*	Degree 2; coef0 1	$\alpha \in \{1, 10, 100\}$; $\gamma \in \{0.3, 1\}/p$
Kernel	SVR, RBF*	—	$C \in \{0.3, 1, 3\}$; $\gamma \in \{0.3, 1\}/p$; $\varepsilon \in \{0.2, 0.5\}$
Kernel	SVR, linear*	—	$C \in \{0.001, 0.01, 0.1\}$
Kernel	Gaussian process	10 principal components; constant × RBF (length scale 3) + white noise; kernel parameters by maximum marginal likelihood (one restart); normalized outcome	—
Nearest neighbors	k-nearest neighbors	10 screened features; distance weights	$k \in \{5, 10, 20\}$
Trees	Random forest	400 trees; minimum leaf 3; max features 0.33	—
Trees	Extremely randomized trees	400 trees; minimum leaf 3; max features 0.5	—
Boosting	Gradient boosting (scikit-learn)	300 trees; learning rate 0.03; depth 2; subsample 0.7; max features 0.5	—
Boosting	Histogram gradient boosting	200 iterations; learning rate 0.05; depth 3; minimum leaf 10; L2 penalty 1	—
Boosting	XGBoost	300 trees; learning rate 0.03; depth 3; subsample 0.7; column sample 0.5; minimum child weight 3; λ 1	—
Boosting	LightGBM	300 trees; learning rate 0.03; 7 leaves; minimum child samples 10; subsample 0.7; column sample 0.5	—
Boosting	CatBoost	400 iterations; depth 4; learning rate 0.05; L2 penalty 3; feature sample 0.5	—
Neural network	Multilayer perceptron*	30 screened features; one hidden layer of 16 units; L-BFGS; 2,000 iterations	α (L2 penalty) $\in \{1, 10, 30\}$

*: outcome standardized within the training fold. k: number of features kept by univariate F-test screening. p: number of numerical features plus 8, used to scale kernel widths. Ridge penalties written as a range were chosen inside the learner (RidgeCV).

In the window-augmented design (Table S13), every well-covered window of a campus is a row carrying the campus's score. Rows were weighted by the inverse of the campus's number of windows in every learner except k-nearest neighbors, inner folds were grouped by campus, and a campus's prediction is the mean of its window predictions weighted by the number of 20 m units observed. Leaves and neighborhoods are larger than in Table S12 because the rows of one campus are not independent.

The wide search (Table S14) drew settings at random from the distributions shown, with a seed fixed for each outer fold, and scored each setting by campus-level 5-fold cross-validation inside the outer training set (Bergstra & Bengio, 2012). It was run on the first random split (seed 1000) and on the state-grouped split. In the window-augmented design, rows were weighted as above wherever a learner accepts sample weights. Median regression and the Gaussian process were fitted with one row per campus only.

Table S13: Learners of the window-augmented design.

Learner	Fixed settings	Tuned in inner cross-validation
Ridge*	—	$\alpha \in \{1, 10, 100, 1000\}$
Elastic net*	—	$\alpha \in \{0.03, 0.1, 0.3\}$; L1 ratio $\in \{0.2, 0.7\}$
Kernel ridge, RBF*	—	$\alpha \in \{1, 10\}$; $\gamma \in \{0.003, 0.01, 0.03\}$
SVR, RBF*	ε 0.5	C $\in \{0.3, 1\}$; $\gamma \in \{0.003, 0.01, 0.03\}$
k-nearest neighbors	Distance weights; rows not weighted	k $\in \{20, 50, 100\}$
Random forest	300 trees; minimum leaf 20; max features 0.33	—
Extremely randomized trees	300 trees; minimum leaf 20; max features 0.5	—
LightGBM	300 trees; learning rate 0.03; 7 leaves; minimum child samples 40; subsample 0.7; column sample 0.5; λ 5	—
CatBoost	400 iterations; depth 4; learning rate 0.05; L2 penalty 10; feature sample 0.5	—

*: outcome standardized within the training fold.

Table S14: Learners and search ranges of the wide search.

Family	Learner	Search distribution	Settings drawn (per campus / window-augmented)
Linear	Ridge*	α: LU[0.001, 10,000]	20 / 12
Linear	Lasso*	α: LU[0.0001, 1]	20 / 12
Linear	Elastic net*	α: LU[0.0001, 3]; L1 ratio: U[0.05, 0.95]	20 / 12
Linear	Bayesian ridge*	Prior parameters $\alpha 1$ and $\lambda 1$: LU[1e−7, 0.001]	20 / 12
Linear	ARD regression*	Pruning threshold: LU[100, 100,000]; 500 iterations	20 / 12
Linear	Huber regression*	ε: U[1.1, 2.0]; α: LU[0.0001, 100]	20 / 12
Linear	Lasso by least-angle regression*	Information criterion $\in$ {AIC, BIC}	2 / 2
Linear	Orthogonal matching pursuit*	Nonzero coefficients: {1, …, 15}	20 / 12
Linear	Partial least squares*	Components: {1, …, 9}	20 / 12
Linear	Principal-component regression*	Components: {2, …, 29}; ridge α: LU[0.001, 1,000]	20 / 12
Linear	Ridge after screening*	k: {3, …, 39}; α: LU[0.001, 1,000]	20 / 12
Linear	Median regression*	α: LU[0.0001, 1]; HiGHS solver	20 / —
Polynomial	Degree-2 ridge*	k: {3, …, 15}; α: LU[0.01, 10,000]	20 / 12
Polynomial	Degree-2 lasso*	k: {3, …, 15}; α: LU[0.001, 1]	20 / 12
Polynomial	Degree-3 ridge*	k: {2, …, 7}; α: LU[0.1, 100,000]	20 / 12
Additive (splines)	Spline ridge*	Knots: {3, …, 6}; degree: {1, 2, 3}; α: LU[0.1, 100,000]	20 / 12
Kernel	Kernel ridge, RBF*	α: LU[0.001, 10]; γ: LU[0.0001, 0.3]	20 / 12
Kernel	Kernel ridge, Laplacian*	α: LU[0.001, 10]; γ: LU[0.0001, 0.3]	20 / 12
Kernel	Kernel ridge, polynomial*	α: LU[0.01, 100]; degree $\in \{2, 3\}$; γ: LU[0.001, 0.1]; coef0: U[0, 2]	20 / 12
Kernel	SVR, RBF*	C: LU[0.05, 100]; γ: LU[0.0001, 0.3]; ε: LU[0.01, 1]	20 / 12

Family	Learner	Search distribution	Settings drawn (per campus / window-augmented)
Kernel	ν-SVR, RBF*	C: LU[0.05, 100]; γ: LU[0.0001, 0.3]; ν: U[0.1, 0.9]	20 / 12
Kernel	SVR, linear*	C: LU[0.0001, 1]; ε: LU[0.01, 1]	20 / 12
Kernel	Gaussian process*	Kernel $\in$ {RBF, Matérn ν = 1.5, Matérn ν = 2.5, rational quadratic, linear + RBF}, each plus white noise, with parameters by maximum marginal likelihood; principal components $\in$ {5, 10, 20}	15 / —
Nearest neighbors	k-nearest neighbors*	k: {3, . . . , 40}; weights $\in$ {uniform, distance}; Minkowski power $\in$ {1, 2}	20 / 12
Trees	Random forest	300 trees; max features: U[0.1, 1.0]; minimum leaf: {1, . . . , 15}; max depth $\in$ {none, 3, 5, 8, 12}	12 / 6
Trees	Extremely randomized trees	As random forest	12 / 6
Boosting	Gradient boosting (scikit-learn)	Learning rate: LU[0.005, 0.2]; trees: {50, . . . , 599}; depth: {1, . . . , 4}; subsample: U[0.5, 1.0]; minimum leaf: {3, . . . , 20}	12 / 6
Boosting	Histogram gradient boosting	Learning rate: LU[0.01, 0.3]; iterations: {50, . . . , 399}; maximum leaf nodes: {4, . . . , 31}; L2 penalty: LU[0.001, 10]; minimum leaf: {5, . . . , 30}	12 / 6
Boosting	XGBoost	Trees: {50, . . . , 599}; learning rate: LU[0.005, 0.3]; depth: {1, . . . , 5}; subsample: U[0.5, 1.0]; column sample: U[0.3, 1.0]; λ: LU[0.001, 100]; minimum child weight: LU[0.5, 20]	12 / 6
Boosting	LightGBM	Trees: {50, . . . , 599}; learning rate: LU[0.005, 0.3]; leaves: {4, . . . , 31}; minimum child samples: {5, . . . , 40}; subsample: U[0.5, 1.0]; column sample: U[0.3, 1.0]; λ: LU[0.001, 100]	12 / 6
Boosting	CatBoost	Iterations: {200, . . . , 799}; depth: {2, . . . , 6}; learning rate: LU[0.01, 0.2]; L2 penalty: LU[1, 30]	8 / 5
Boosting	AdaBoost	Regression trees of depth {1, . . . , 4}; estimators: {50, . . . , 399}; learning rate: LU[0.01, 1]; loss $\in$ {linear, square}	12 / 6
Neural network	Multilayer perceptron*	Hidden layers $\in$ {16, 32, 64, 32–16, 64–32 units}; α (L2 penalty): LU[0.0001, 10]; initial learning rate: LU[0.0001, 0.01]; Adam; 3,000 iterations	8 / 5

LU[a, b]: log-uniform on [a, b]; U[a, b]: uniform on [a, b]; {a, . . . , b}: integers from a to b. *: outcome standardized within the training fold. k: number of features kept by univariate F-test screening. —: learner not used in that design. The least-angle lasso has only two settings. A run evaluated about 5,500 settings across its outer folds and two splits (about 3,000 in the window-augmented design).

Note S4. Analysis campuses

Table S15 lists the 252 analysis campuses, in 38 states and the District of Columbia. The sample column shows why each campus is included. In all, 198 are the main campuses of institutions in the U.S. News 2026 top 100 national universities and liberal arts colleges (Top 100), 41 others have a STARS report (STARS), and 13 others have a QS Sustainability score (QS); campuses in the first group may also have STARS or QS scores. Altogether, 130 campuses have a STARS overall score (92 under version 2.2 and 38 under version 3.0), 109 of them with public sub-scores, and 113 have a QS Sustainability score. Three campuses (Colorado Mountain College, Principia College, and the University of Minnesota-Rochester) have no usable primary window, so the models use 128 STARS

campuses and 113 QS campuses, and the index covers 249.

Table S15: The 252 analysis campuses.

Campus	State	Carnegie	Locale	Sample	STARS overall (version)	Sub-scores	QS
Auburn University	AL	R1	Urban	QS	—	—	10.9
Hendrix College	AR	Bacc	Urban	Top 100	—	—	—
California Institute of Technology	CA	R1	Urban	Top 100	—	—	62.5
California State University-Fullerton	CA	R2	Suburban	STARS	71.02 (2.2)	yes	—
California State University-Long Beach	CA	R2	Urban	STARS	63.01 (2.2)	yes	1.0
Claremont McKenna College	CA	Bacc	Suburban	Top 100	40.45 (3.0)	yes	—
Harvey Mudd College	CA	Bacc	Suburban	Top 100	—	—	—
Loyola Marymount University	CA	R2	Urban	STARS	72.75 (3.0)	yes	—
Occidental College	CA	Bacc	Urban	Top 100	—	—	—
Pepperdine University	CA	D/PU	Suburban	Top 100	—	—	—
Pitzer College	CA	Bacc	Suburban	Top 100	65.39 (2.2)	no	—
Pomona College	CA	Bacc	Suburban	Top 100	—	—	—
Saint Mary's College of California	CA	—	—	STARS	67.92 (2.2)	no	—
San Diego State University	CA	R2	Urban	STARS	78.11 (2.2)	yes	9.4
Santa Clara University	CA	D/PU	Urban	Top 100	—	—	—
Scripps College	CA	Bacc	Suburban	Top 100	—	—	—
Soka University of America	CA	Bacc	Suburban	Top 100	51.84 (3.0)	yes	—
Stanford University	CA	R1	Suburban	Top 100	87.04 (3.0)	yes	81.2
University of California-Berkeley	CA	R1	Urban	Top 100	86.76 (3.0)	yes	100.0
University of California-Davis	CA	R1	Suburban	Top 100	79.54 (2.2)	no	76.1
University of California-Irvine	CA	R1	Urban	Top 100	84.49 (2.2)	yes	36.0
University of California-Los Angeles	CA	R1	Urban	Top 100	71.84 (2.2)	yes	67.1
University of California-Merced	CA	R2	Suburban	Top 100	—	—	—
University of California-Riverside	CA	R1	Urban	Top 100	74.80 (3.0)	yes	21.3
University of California-San Diego	CA	R1	Urban	Top 100	86.87 (2.2)	yes	44.3
University of California-Santa Barbara	CA	R1	Suburban	Top 100	78.00 (2.2)	yes	28.4
University of California-Santa Cruz	CA	R1	Urban	Top 100	77.70 (3.0)	yes	83.6
University of San Diego	CA	R2	Urban	STARS	65.07 (2.2)	yes	1.0
University of San Francisco	CA	D/PU	Urban	QS	—	—	1.2
University of Southern California	CA	R1	Urban	Top 100	69.91 (2.2)	yes	97.7
University of the Pacific	CA	D/PU	Urban	STARS	58.33 (2.2)	yes	1.0
Whittier College	CA	Bacc	Suburban	Top 100	—	—	—
Colorado College	CO	Bacc	Urban	Top 100	65.41 (2.2)	yes	—
Colorado Mountain College	CO	—	—	STARS	48.69 (2.2)	yes	—
Colorado School of Mines	CO	R1	Suburban	Top 100	66.82 (2.2)	yes	18.0
United States Air Force Academy	CO	—	—	Top 100	—	—	—
University of Colorado Boulder	CO	R1	Urban	Top 100	68.18 (3.0)	yes	62.8
University of Denver	CO	R1	Urban	STARS	46.05 (2.2)	yes	9.9
Connecticut College	CT	Bacc	Urban	Top 100	71.19 (2.2)	yes	—
Trinity College	CT	Bacc	Urban	Top 100	52.60 (2.2)	yes	—
University of Connecticut	CT	R1	Town	Top 100	84.74 (2.2)	yes	37.6
Wesleyan University	CT	Bacc	Urban	Top 100	65.69 (2.2)	no	—
Yale University	CT	R1	Urban	Top 100	—	—	92.2
George Washington University	DC	R1	Urban	Top 100	85.19 (2.2)	yes	27.3

Campus	State	Carnegie	Locale	Sample	STARS overall (version)	Sub-scores	QS
Georgetown University	DC	R1	Urban	Top 100	—	—	37.0
Howard University	DC	R2	Urban	Top 100	—	—	1.1
University of Delaware	DE	R1	Suburban	Top 100	49.58 (2.2)	yes	65.3
Florida International University	FL	R1	Suburban	Top 100	66.01 (2.2)	yes	81.2
Florida State University	FL	R1	Urban	Top 100	67.49 (2.2)	no	26.8
University of Central Florida	FL	R1	Suburban	STARS	56.79 (2.2)	yes	12.1
University of Florida	FL	R1	Urban	Top 100	—	—	60.6
University of Miami	FL	R1	Suburban	Top 100	82.69 (2.2)	yes	27.9
University of South Florida	FL	R1	Urban	QS	—	—	15.9
Agnes Scott College	GA	Bacc	Suburban	Top 100	66.19 (3.0)	yes	—
Emory University	GA	R1	Urban	Top 100	72.29 (2.2)	yes	37.0
Georgia Institute of Technology-Main Campus	GA	R1	Urban	Top 100	65.15 (2.2)	yes	39.5
Morehouse College	GA	Bacc	Urban	Top 100	—	—	—
Spelman College	GA	Bacc	Urban	Top 100	—	—	—
University of Georgia	GA	R1	Urban	Top 100	—	—	30.7
University of North Georgia	GA	—	—	STARS	37.09 (2.2)	yes	—
Grinnell College	IA	Bacc	Town	Top 100	—	—	—
Iowa State University	IA	R1	Urban	QS	—	—	31.3
University of Iowa	IA	R1	Urban	STARS	65.91 (2.2)	yes	20.8
Augustana College	IL	Bacc	Urban	Top 100	—	—	—
Illinois Institute of Technology	IL	R2	Urban	QS	—	—	1.6
Knox College	IL	Bacc	Town	Top 100	45.14 (2.2)	yes	—
Lake Forest College	IL	Bacc	Suburban	Top 100	—	—	—
Loyola University Chicago	IL	R2	Urban	STARS	82.81 (2.2)	no	1.9
Northwestern University	IL	R1	Urban	Top 100	—	—	55.4
Principia College	IL	Bacc	Rural	Top 100	—	—	—
University of Chicago	IL	R1	Urban	Top 100	—	—	46.9
University of Illinois Chicago	IL	R1	Urban	STARS	60.86 (3.0)	yes	—
University of Illinois Urbana-Champaign	IL	R1	Urban	Top 100	73.56 (2.2)	yes	—
Wheaton College	IL	Bacc	Suburban	Top 100	—	—	—
DePauw University	IN	Bacc	Town	Top 100	—	—	—
Earlham College	IN	Bacc	Town	Top 100	—	—	—
Indiana University-Bloomington	IN	R1	Urban	Top 100	71.55 (2.2)	yes	32.7
Indiana University-Purdue University-Indianapolis	IN	R2	Urban	QS	—	—	6.7
Purdue University-Main Campus	IN	R1	Urban	Top 100	54.05 (2.2)	no	90.1
Saint Mary's College	IN	Bacc	Suburban	Top 100	—	—	—
University of Notre Dame	IN	R1	Suburban	Top 100	49.83 (3.0)	yes	68.9
Wabash College	IN	Bacc	Town	Top 100	—	—	—
University of Kansas	KS	R1	Urban	QS	—	—	32.4
Berea College	KY	Bacc	Town	Top 100	—	—	—
Centre College	KY	Bacc	Town	Top 100	—	—	—
University of Kentucky	KY	R1	Urban	STARS	49.19 (3.0)	yes	10.1
Tulane University of Louisiana	LA	R1	Urban	Top 100	54.25 (3.0)	yes	—
Amherst College	MA	Bacc	Urban	Top 100	67.18 (3.0)	yes	—
Boston College	MA	R1	Urban	Top 100	—	—	7.1
Boston University	MA	R1	Urban	Top 100	65.38 (3.0)	yes	65.9
Brandeis University	MA	R1	Urban	Top 100	—	—	1.1
Clark University	MA	R2	Urban	QS	—	—	6.5
College of the Holy Cross	MA	Bacc	Urban	Top 100	54.32 (3.0)	yes	—
Harvard University	MA	R1	Urban	Top 100	—	—	84.4
Massachusetts Institute of Technology	MA	R1	Urban	Top 100	69.23 (3.0)	yes	99.0
Mount Holyoke College	MA	Bacc	Suburban	Top 100	—	—	—

Campus	State	Carnegie	Locale	Sample	STARS overall (version)	Sub-scores	QS
Northeastern University	MA	R1	Urban	Top 100	61.60 (2.2)	no	72.3
Smith College	MA	Bacc	Urban	Top 100	—	—	—
Stonehill College	MA	Bacc	Suburban	Top 100	—	—	—
Tufts University	MA	R1	Suburban	Top 100	—	—	81.0
University of Massachusetts-Amherst	MA	R1	Urban	Top 100	75.29 (2.2)	no	83.4
University of Massachusetts-Boston	MA	R2	Urban	STARS	48.58 (2.2)	yes	8.4
Wellesley College	MA	Bacc	Suburban	Top 100	—	—	—
Wheaton College	MA	Bacc	Suburban	Top 100	—	—	—
Williams College	MA	Bacc	Town	Top 100	68.27 (2.2)	yes	—
Worcester Polytechnic Institute	MA	R2	Urban	Top 100	—	—	1.9
Johns Hopkins University	MD	R1	Urban	Top 100	—	—	61.9
St. John's College	MD	Bacc	Suburban	Top 100	—	—	—
St. Mary's College of Maryland	MD	Bacc	Rural	Top 100	—	—	—
United States Naval Academy	MD	Bacc	Suburban	Top 100	—	—	—
University of Maryland, Baltimore	MD	—	—	STARS	60.36 (3.0)	yes	—
University of Maryland-Baltimore County	MD	R1	Suburban	STARS	65.51 (3.0)	yes	57.7
University of Maryland-College Park	MD	R1	Suburban	Top 100	68.27 (3.0)	yes	54.4
Washington College	MD	Bacc	Town	Top 100	—	—	—
Bates College	ME	Bacc	Urban	Top 100	72.89 (2.2)	yes	—
Bowdoin College	ME	Bacc	Town	Top 100	68.31 (2.2)	yes	—
Colby College	ME	Bacc	Town	Top 100	—	—	—
Hillsdale College	MI	Bacc	Town	Top 100	—	—	—
Hope College	MI	Bacc	Urban	Top 100	65.28 (2.2)	no	—
Kalamazoo College	MI	Bacc	Urban	Top 100	—	—	—
Michigan State University	MI	R1	Urban	Top 100	—	—	96.3
University of Michigan-Ann Arbor	MI	R1	Urban	Top 100	—	—	62.2
Carleton College	MN	Bacc	Town	Top 100	—	—	—
Gustavus Adolphus College	MN	Bacc	Town	Top 100	65.30 (3.0)	yes	—
Macalester College	MN	Bacc	Urban	Top 100	80.87 (2.2)	yes	—
St Olaf College	MN	Bacc	Town	Top 100	—	—	—
University of Minnesota-Crookston	MN	—	—	STARS	65.19 (2.2)	yes	—
University of Minnesota-Duluth	MN	—	—	STARS	72.29 (2.2)	yes	—
University of Minnesota-Rochester	MN	—	—	STARS	56.58 (3.0)	yes	—
University of Minnesota-Twin Cities	MN	R1	Urban	Top 100	65.09 (2.2)	no	96.5
Saint Louis University	MO	R2	Urban	QS	—	—	1.7
University of Missouri-Columbia	MO	R1	Urban	QS	—	—	12.3
University of Missouri-Kansas City	MO	R2	Urban	QS	—	—	1.8
Washington University in St Louis	MO	R1	Suburban	Top 100	66.94 (3.0)	yes	86.7
Davidson College	NC	Bacc	Suburban	Top 100	—	—	—
Duke University	NC	R1	Urban	Top 100	—	—	75.0
Elon University	NC	D/PU	Suburban	STARS	59.20 (2.2)	yes	—
North Carolina State University at Raleigh	NC	R1	Urban	Top 100	65.62 (3.0)	yes	—
University of North Carolina at Chapel Hill	NC	R1	Urban	Top 100	67.20 (3.0)	yes	—

Campus	State	Carnegie	Locale	Sample	STARS overall (version)	Sub-scores	QS
University of North Carolina at Charlotte	NC	R2	Urban	STARS	66.23 (2.2)	yes	4.2
Wake Forest University	NC	R2	Urban	Top 100	—	—	1.7
Central Community College	NE	—	—	STARS	67.45 (2.2)	no	—
Creighton University	NE	R2	Urban	STARS	53.05 (2.2)	no	—
Dartmouth College	NH	R1	Town	Top 100	—	—	71.2
Saint Anselm College	NH	Bacc	Suburban	Top 100	—	—	—
University of New Hampshire-Main Campus	NH	R1	Town	STARS	85.89 (2.2)	yes	5.5
Drew University	NJ	Bacc	Suburban	Top 100	—	—	—
New Jersey Institute of Technology	NJ	R1	Urban	Top 100	55.05 (2.2)	no	26.5
Princeton University	NJ	R1	Urban	Top 100	72.36 (2.2)	yes	51.5
Rutgers University-Camden	NJ	R2	Urban	Top 100	—	—	—
Rutgers University-New Brunswick	NJ	R1	Urban	Top 100	—	—	95.4
Rutgers University-Newark	NJ	R2	Urban	Top 100	—	—	29.3
Stevens Institute of Technology	NJ	R2	Suburban	Top 100	66.99 (2.2)	yes	1.3
St. John's College	NM	Bacc	Rural	Top 100	—	—	—
Bard College	NY	Bacc	Rural	Top 100	65.04 (2.2)	yes	—
Barnard College	NY	Bacc	Urban	Top 100	69.22 (2.2)	yes	—
Binghamton University	NY	R1	Suburban	Top 100	73.80 (2.2)	no	—
Clarkson University	NY	R2	Town	STARS	68.91 (2.2)	yes	1.2
Colgate University	NY	Bacc	Rural	Top 100	—	—	—
Columbia University in the City of New York	NY	R1	Urban	Top 100	—	—	—
Cornell University	NY	R1	Urban	Top 100	86.04 (2.2)	yes	72.1
Fordham University	NY	R2	Urban	Top 100	41.82 (3.0)	yes	1.4
Hamilton College	NY	Bacc	Suburban	Top 100	—	—	—
Hobart William Smith Colleges	NY	Bacc	Town	Top 100	52.04 (3.0)	yes	—
New York University	NY	R1	Urban	Top 100	72.23 (2.2)	yes	95.4
Rensselaer Polytechnic Institute	NY	R1	Urban	Top 100	—	—	2.8
Rochester Institute of Technology	NY	R2	Suburban	Top 100	—	—	34.5
Skidmore College	NY	Bacc	Suburban	Top 100	51.09 (3.0)	yes	—
St Lawrence University	NY	Bacc	Town	Top 100	—	—	—
State University of New York at New Paltz	NY	—	—	STARS	65.17 (2.2)	no	—
Stony Brook University	NY	R1	Suburban	Top 100	54.94 (3.0)	yes	—
Syracuse University	NY	R1	Urban	Top 100	64.57 (2.2)	yes	11.6
Union College	NY	Bacc	Urban	Top 100	—	—	—
United States Military Academy	NY	Bacc	Town	Top 100	—	—	—
University of Rochester	NY	R1	Urban	Top 100	45.82 (3.0)	yes	63.7
Vassar College	NY	Bacc	Suburban	Top 100	—	—	—
Yeshiva University	NY	D/PU	Urban	Top 100	—	—	15.8
Case Western Reserve University	OH	R1	Urban	Top 100	—	—	53.1
Denison University	OH	Bacc	Suburban	Top 100	—	—	—
Kenyon College	OH	Bacc	Rural	Top 100	—	—	—
Oberlin College	OH	Bacc	Town	Top 100	74.88 (2.2)	yes	—
The College of Wooster	OH	Bacc	Town	Top 100	—	—	—
University of Dayton	OH	R2	Urban	STARS	65.97 (3.0)	yes	—
University of Central Oklahoma	OK	—	—	QS	—	—	1.0
Lane Community College	OR	—	—	STARS	54.62 (2.2)	yes	—
Lewis & Clark College	OR	Bacc	Urban	Top 100	—	—	—
Oregon State University	OR	R1	Urban	STARS	66.53 (3.0)	yes	21.1
Reed College	OR	Bacc	Urban	Top 100	—	—	—

Campus	State	Carnegie	Locale	Sample	STARS overall (version)	Sub-scores	QS
University of Oregon	OR	R1	Urban	STARS	71.15 (2.2)	yes	16.1
Willamette University	OR	Bacc	Urban	Top 100	—	—	—
Allegheny College	PA	Bacc	Town	Top 100	—	—	—
Bryn Mawr College	PA	Bacc	Suburban	Top 100	45.28 (2.2)	yes	—
Bucknell University	PA	Bacc	Town	Top 100	66.75 (2.2)	no	—
Carnegie Mellon University	PA	R1	Urban	Top 100	75.88 (2.2)	yes	24.5
Dickinson College	PA	Bacc	Urban	Top 100	81.96 (2.2)	yes	—
Drexel University	PA	R1	Urban	Top 100	45.10 (2.2)	yes	8.3
Franklin and Marshall College	PA	Bacc	Urban	Top 100	—	—	—
Gettysburg College	PA	Bacc	Town	Top 100	—	—	—
Haverford College	PA	Bacc	Suburban	Top 100	59.85 (2.2)	yes	—
Juniata College	PA	Bacc	Town	Top 100	47.16 (2.2)	yes	—
Lafayette College	PA	Bacc	Suburban	Top 100	69.35 (2.2)	yes	—
Lehigh University	PA	R2	Urban	Top 100	67.16 (3.0)	yes	33.3
Lycoming College	PA	Bacc	Urban	Top 100	—	—	—
Muhlenberg College	PA	Bacc	Urban	Top 100	68.89 (2.2)	yes	—
Susquehanna University	PA	Bacc	Town	Top 100	—	—	—
Swarthmore College	PA	Bacc	Suburban	Top 100	66.09 (2.2)	no	1.0
Temple University	PA	R1	Urban	STARS	66.57 (2.2)	yes	61.9
University of Pennsylvania	PA	R1	Urban	Top 100	68.42 (2.2)	yes	93.0
University of Pittsburgh-Pittsburgh Campus	PA	R1	Urban	Top 100	72.67 (2.2)	yes	—
Ursinus College	PA	Bacc	Suburban	Top 100	—	—	—
Villanova University	PA	R2	Suburban	Top 100	—	—	—
Washington & Jefferson College	PA	Bacc	Suburban	Top 100	—	—	—
West Chester University of Pennsylvania	PA	R2	Suburban	STARS	56.22 (3.0)	yes	—
Brown University	RI	R1	Urban	Top 100	—	—	46.6
Clemson University	SC	R1	Suburban	Top 100	61.01 (2.2)	yes	12.3
Furman University	SC	Bacc	Suburban	Top 100	69.74 (3.0)	yes	—
Wofford College	SC	Bacc	Urban	Top 100	—	—	—
Rhodes College	TN	Bacc	Urban	Top 100	—	—	—
The University of Tennessee-Knoxville	TN	R1	Urban	QS	—	—	15.0
The University of the South	TN	Bacc	Rural	Top 100	—	—	—
Vanderbilt University	TN	R1	Urban	Top 100	—	—	24.5
Austin College	TX	Bacc	Urban	Top 100	—	—	—
Baylor University	TX	R1	Urban	Top 100	—	—	4.0
Rice University	TX	R1	Urban	Top 100	—	—	20.6
Southern Methodist University	TX	R2	Suburban	Top 100	—	—	2.4
Southwestern University	TX	Bacc	Urban	Top 100	—	—	—
Texas A & M University-College Station	TX	R1	Urban	Top 100	—	—	—
Texas Christian University	TX	R2	Urban	Top 100	—	—	—
The University of Texas at Austin	TX	R1	Urban	Top 100	67.80 (2.2)	yes	52.1
Trinity University	TX	Bacc	Urban	Top 100	—	—	—
University of Houston	TX	R1	Urban	STARS	55.67 (2.2)	yes	19.7
George Mason University	VA	R1	Suburban	STARS	73.66 (2.2)	yes	22.5
Randolph-Macon College	VA	Bacc	Suburban	Top 100	—	—	—
University of Richmond	VA	Bacc	Urban	Top 100	67.31 (2.2)	yes	—
University of Virginia-Main Campus	VA	R1	Suburban	Top 100	74.85 (2.2)	yes	35.4
Virginia Commonwealth University	VA	R1	Urban	STARS	65.54 (2.2)	yes	8.5
Virginia Military Institute	VA	Bacc	Town	Top 100	—	—	—

Campus	State	Carnegie	Locale	Sample	STARS overall (version)	Sub-scores	QS
Virginia Polytechnic Institute and State University	VA	R1	Urban	Top 100	—	—	34.8
Virginia Wesleyan University	VA	Bacc	Urban	STARS	53.11 (2.2)	no	—
Washington and Lee University	VA	Bacc	Town	Top 100	60.54 (2.2)	yes	—
William & Mary	VA	R2	Suburban	Top 100	—	—	—
Middlebury College	VT	Bacc	Town	Top 100	—	—	—
University of Vermont	VT	R2	Urban	STARS	76.71 (2.2)	no	8.1
Gonzaga University	WA	D/PU	Urban	STARS	65.01 (2.2)	yes	—
University of Puget Sound	WA	Bacc	Urban	Top 100	45.33 (3.0)	yes	—
University of Washington-Bothell Campus	WA	—	—	STARS	70.01 (3.0)	yes	—
University of Washington-Seattle Campus	WA	R1	Urban	Top 100	67.89 (3.0)	yes	—
Whitman College	WA	Bacc	Urban	Top 100	58.23 (2.2)	no	—
Beloit College	WI	Bacc	Urban	Top 100	—	—	—
Lawrence University	WI	Bacc	Urban	Top 100	—	—	—
Marquette University	WI	R2	Urban	Top 100	—	—	1.2
University of Wisconsin-Madison	WI	R1	Urban	Top 100	67.70 (2.2)	yes	59.6

Campus names follow the HIFLD campus layer; campuses are ordered by state and name. Carnegie: R1 and R2, doctoral universities with very high and high research activity; D/PU, doctoral universities with moderate research activity; Bacc, baccalaureate colleges with an arts and sciences focus. Carnegie class and locale come from the project's IPEDS-derived file. STARS overall: the score on the dated report page, with the STARS version in parentheses. Sub-scores: whether the report page shows subcategory points. QS: Sustainability lens score, QS World University Rankings 2025. —: not available. The same list with IPEDS identifiers is in tableS15_campuses.csv.

Note S5. Visible Campus Environment Index by campus

Table S16 reports the Visible Campus Environment Index and its rank for each of the 249 campuses with a usable primary window, with the 90% rank interval, the three dimension scores, and the index under the alternative specifications compared in Section 5.7.

Table S16: Visible Campus Environment Index and rank of each campus.

						Dimension scores			Alternative specifications		
Rank	Campus	State	Primary window	VCEI	90% rank interval	Green	Pedes.	Low car	Arith.	Ver. A	Leaf-on
1	Wellesley College	MA	Sep–Oct 2012	0.90	1–15	1.00	0.75	0.97	0.91	0.67	0.63
2	Denison University	OH	Jul 2007	0.89	1–10	0.83	0.92	0.91	0.89	0.59	0.89
3	Agnes Scott College	GA	Jul 2014	0.85	2–20	0.95	0.73	0.88	0.85	0.57	0.85
4	Augustana College	IL	Aug–Sep 2011	0.81	1–33	0.81	1.00	0.67	0.82	0.55	0.81
5	Colgate University	NY	Jun 2011	0.78	2–38	0.81	0.60	0.99	0.80	0.51	0.78
6	Colby College	ME	Aug 2007	0.77	5–34	0.69	0.83	0.81	0.78	0.54	0.77
7	Lake Forest College	IL	Jul 2007	0.76	2–60	0.58	1.00	0.74	0.77	0.56	0.76
8	St. John's College	MD	Aug 2021	0.74	7–34	0.74	0.73	0.75	0.74	0.72	0.74
9	Lawrence University	WI	Sep–Oct 2012	0.73	3–65	0.58	1.00	0.68	0.75	0.56	0.78
10	Spelman College	GA	May–Jun 2016	0.70	5–75	0.56	1.00	0.62	0.73	0.51	0.70
11	Earlham College	IN	May 2019	0.70	12–49	0.67	0.73	0.71	0.70	0.68	0.70
12	Ursinus College	PA	Oct–Nov 2017	0.70	10–71	0.55	0.87	0.72	0.71	0.52	0.90

Rank	Campus	State	Primary window	VCEI	90% rank interval	Dimension scores			Alternative specifications		
						Green	Pedes.	Low car	Arith.	Ver. A	Leaf-on
13	University of Richmond	VA	Aug–Sep 2007	0.69	13–47	0.73	0.60	0.75	0.70	0.46	0.69
14	Boston College	MA	Jul–Aug 2017	0.69	13–61	0.65	0.79	0.64	0.69	0.45	0.69
15	University of Wisconsin-Madison	WI	Aug 2018	0.69	15–54	0.67	0.72	0.68	0.69	0.50	0.69
16	Swarthmore College	PA	Sep–Oct 2007	0.68	9–69	0.83	0.50	0.76	0.70	0.48	0.50
17	Brown University	RI	Aug–Sep 2017	0.68	6–85	0.55	1.00	0.58	0.71	0.48	0.68
18	Wofford College	SC	Jul 2019	0.68	12–60	0.79	0.55	0.70	0.68	0.48	0.68
19	Lehigh University	PA	Jul 2011	0.67	4–116	0.95	0.38	0.85	0.73	0.38	0.67
20	Lafayette College	PA	Sep 2017	0.67	13–69	0.79	0.53	0.73	0.68	0.41	0.67
21	Skidmore College	NY	Oct 2007	0.66	11–79	0.84	0.48	0.73	0.68	0.44	0.20
22	St Olaf College	MN	Jun 2019	0.66	10–102	0.81	0.43	0.83	0.69	0.67	0.66
23	Wheaton College	IL	Jun–Jul 2012	0.65	13–88	0.54	0.86	0.59	0.67	0.45	0.65
24	Carleton College	MN	Aug–Sep 2007	0.65	23–64	0.62	0.60	0.72	0.65	0.46	0.65
25	Yale University	CT	Jun–Jul 2011	0.64	10–111	0.60	0.92	0.48	0.66	0.42	0.64
26	Princeton University	NJ	Aug–Sep 2017	0.64	19–101	0.78	0.44	0.76	0.66	0.39	0.64
27	Kalamazoo College	MI	Jun–Jul 2012	0.64	24–68	0.67	0.63	0.61	0.64	0.39	0.64
28	Beloit College	WI	Oct 2007	0.64	18–87	0.58	0.79	0.57	0.64	0.47	0.50
29	Trinity University	TX	Jul 2008	0.63	6–131	0.36	1.00	0.68	0.68	0.47	0.63
30	Duke University	NC	Aug 2011	0.62	13–101	0.85	0.46	0.61	0.64	0.43	0.62
31	Allegheny College	PA	Jul–Aug 2014	0.62	22–104	0.68	0.44	0.79	0.64	0.41	0.62
32	Wabash College	IN	Sep 2013	0.62	21–97	0.60	0.75	0.53	0.62	0.46	0.62
33	Washington and Lee University	VA	Aug 2012	0.61	23–114	0.79	0.42	0.70	0.64	0.39	0.61
34	Oberlin College	OH	Oct 2013	0.61	15–120	0.69	0.40	0.83	0.64	0.36	0.60
35	Wake Forest University	NC	Sep 2014	0.61	17–111	0.85	0.44	0.60	0.63	0.38	0.61
36	St. John's College	NM	Dec 2007	0.60	5–149	0.31	0.79	0.91	0.67	0.40	0.55
37	The University of the South	TN	Dec 2007	0.60	14–102	0.48	0.54	0.84	0.62	0.40	0.25
38	Amherst College	MA	Oct 2017	0.60	28–117	0.75	0.41	0.70	0.62	0.36	0.52
39	University of Missouri-Columbia	MO	Jun–Jul 2017	0.60	19–113	0.50	0.84	0.50	0.62	0.44	0.60
40	University of Georgia	GA	May 2012	0.59	31–92	0.70	0.55	0.54	0.60	0.40	0.59
41	The College of Wooster	OH	Aug 2012	0.59	41–97	0.66	0.48	0.64	0.60	0.38	0.59
42	Wesleyan University	CT	Oct–Nov 2020	0.59	35–94	0.57	0.49	0.72	0.60	0.59	0.66
43	Cornell University	NY	Jul–Aug 2018	0.59	32–123	0.69	0.40	0.73	0.61	0.36	0.59
44	Muhlenberg College	PA	Jun–Jul 2011	0.59	37–82	0.60	0.56	0.60	0.59	0.34	0.59
45	Kenyon College	OH	May–Jun 2012	0.59	6–162	0.90	0.25	0.90	0.68	0.29	0.59
46	Reed College	OR	Aug–Sep 2011	0.59	10–152	0.91	0.29	0.77	0.66	0.32	0.59
47	Saint Mary's College	IN	May–Jun 2023	0.58	17–142	0.70	0.34	0.83	0.62	0.56	0.58
48	Haverford College	PA	Sep–Oct 2012	0.58	8–165	0.91	0.24	0.85	0.67	0.19	0.47
49	Virginia Wesleyan University	VA	Sep–Oct 2007	0.57	41–96	0.55	0.50	0.69	0.58	0.41	0.51
50	Whitman College	WA	Jul 2018	0.57	44–89	0.60	0.55	0.56	0.57	0.38	0.57
51	Tulane University of Louisiana	LA	May 2012	0.57	37–103	0.52	0.66	0.55	0.58	0.33	0.57
52	Elon University	NC	Aug 2013	0.57	41–119	0.61	0.43	0.71	0.58	0.38	0.57
53	Dartmouth College	NH	Oct 2010	0.56	46–103	0.49	0.58	0.62	0.56	0.32	0.59
54	Rice University	TX	Dec 2018–Jan 2019	0.56	50–93	0.56	0.56	0.56	0.56	0.54	0.62
55	Drew University	NJ	Sep 2013	0.56	4–180	1.00	0.21	0.84	0.68	0.21	0.56
56	Saint Anselm College	NH	Aug 2011	0.56	11–171	0.85	0.24	0.86	0.65	0.24	0.56
57	Bates College	ME	Aug 2009	0.56	20–153	0.32	0.88	0.60	0.60	0.35	0.56
58	Davidson College	NC	Sep–Oct 2012	0.55	26–145	0.83	0.35	0.59	0.59	0.33	0.49
59	Southern Methodist University	TX	May–Jun 2014	0.55	25–138	0.43	0.83	0.47	0.58	0.37	0.55

Rank	Campus	State	Primary window	VCEI	90% rank interval	Dimension scores			Alternative specifications		
						Green	Pedes.	Low car	Arith.	Ver. A	Leaf-on
60	Indiana University-Bloomington	IN	Jul 2015	0.55	56–120	0.63	0.45	0.53	0.55	0.35	0.55
61	William & Mary	VA	Oct–Nov 2007	0.55	42–126	0.68	0.51	0.47	0.56	0.39	0.74
62	Whittier College	CA	Jul 2007	0.55	17–162	0.29	0.89	0.64	0.61	0.40	0.55
63	Union College	NY	Sep 2015	0.55	46–128	0.63	0.55	0.47	0.55	0.35	0.55
64	Knox College	IL	Aug 2011	0.55	36–124	0.47	0.71	0.49	0.56	0.40	0.55
65	University of Florida	FL	Nov–Dec 2007	0.54	56–114	0.48	0.56	0.59	0.54	0.40	0.55
66	University of Washington-Seattle Campus	WA	Aug–Sep 2017	0.54	58–113	0.54	0.56	0.52	0.54	0.34	0.54
67	Grinnell College	IA	Oct 2007	0.53	36–146	0.34	0.70	0.64	0.56	0.38	0.58
68	University of North Carolina at Chapel Hill	NC	Aug–Sep 2012	0.53	40–153	0.74	0.32	0.64	0.57	0.31	0.53
69	Oregon State University	OR	Sep 2011	0.53	47–126	0.44	0.67	0.50	0.54	0.36	0.53
70	Wheaton College	MA	Sep–Oct 2012	0.53	45–154	0.70	0.31	0.67	0.56	0.29	0.49
71	Iowa State University	IA	Aug–Sep 2019	0.53	69–109	0.50	0.50	0.59	0.53	0.52	0.53
72	Worcester Polytechnic Institute	MA	Aug–Sep 2007	0.52	30–160	0.45	0.80	0.39	0.55	0.36	0.52
73	Clemson University	SC	Jun 2019	0.52	54–139	0.64	0.48	0.46	0.53	0.37	0.52
74	Austin College	TX	Dec 2007	0.52	8–184	0.23	1.00	0.61	0.61	0.46	0.55
75	Randolph-Macon College	VA	May 2023	0.52	34–150	0.78	0.39	0.45	0.54	0.49	0.52
76	Trinity College	CT	Jul 2015	0.51	62–144	0.59	0.51	0.45	0.52	0.33	0.51
77	Baylor University	TX	Oct 2015	0.51	10–170	0.31	1.00	0.44	0.58	0.39	0.54
78	St Lawrence University	NY	Aug 2011	0.51	44–162	0.62	0.30	0.73	0.55	0.29	0.51
79	Case Western Reserve University	OH	May–Jun 2019	0.51	27–159	0.35	0.86	0.44	0.55	0.50	0.51
80	Rutgers University-New Brunswick	NJ	Sep–Oct 2013	0.50	67–142	0.61	0.41	0.52	0.51	0.35	0.39
81	University of Miami	FL	Feb–Mar 2019	0.50	78–122	0.51	0.49	0.51	0.50	0.51	0.54
82	Tufts University	MA	Aug 2011	0.50	77–130	0.49	0.53	0.49	0.50	0.30	0.50
83	Rhodes College	TN	Jun 2019	0.50	49–162	0.72	0.31	0.56	0.53	0.51	0.50
84	University of Connecticut	CT	Nov 2017	0.50	49–166	0.31	0.65	0.61	0.53	0.36	0.52
85	University of Kansas	KS	May 2011	0.50	77–132	0.47	0.44	0.59	0.50	0.31	0.50
86	University of Minnesota-Duluth	MN	Aug–Sep 2011	0.49	85–122	0.47	0.49	0.52	0.50	0.32	0.49
87	Smith College	MA	Aug 2013	0.49	84–128	0.50	0.45	0.53	0.49	0.31	0.49
88	University of Illinois Urbana-Champaign	IL	Sep–Oct 2015	0.49	65–140	0.39	0.62	0.49	0.50	0.35	0.50
89	Emory University	GA	Apr–May 2012	0.49	29–181	0.84	0.24	0.60	0.56	0.25	0.65
90	Gettysburg College	PA	Jul–Aug 2012	0.49	83–133	0.43	0.48	0.56	0.49	0.34	0.49
91	Susquehanna University	PA	Aug–Sep 2012	0.49	37–169	0.49	0.30	0.78	0.52	0.25	0.49
92	Brandeis University	MA	Oct 2007	0.48	73–147	0.38	0.59	0.50	0.49	0.35	0.40
93	University of Colorado Boulder	CO	May 2011	0.48	72–150	0.37	0.50	0.61	0.49	0.28	0.48
94	Connecticut College	CT	Jul–Aug 2019	0.48	7–202	0.98	0.15	0.74	0.62	0.50	0.48
95	DePauw University	IN	Sep 2007	0.48	11–192	0.20	1.00	0.54	0.58	0.37	0.48
96	Centre College	KY	Jun 2012	0.48	93–137	0.42	0.48	0.53	0.48	0.32	0.48
97	University of Maryland-College Park	MD	Jun–Jul 2008	0.48	91–142	0.45	0.43	0.56	0.48	0.31	0.48
98	North Carolina State University at Raleigh	NC	Dec 2017	0.47	87–150	0.48	0.46	0.49	0.47	0.34	0.47
99	Texas Christian University	TX	Jun–Jul 2018	0.47	34–184	0.38	0.81	0.34	0.51	0.44	0.47
100	Hendrix College	AR	Aug–Sep 2019	0.47	90–148	0.51	0.39	0.53	0.48	0.47	0.47
101	Santa Clara University	CA	Dec 2017	0.47	40–184	0.24	0.68	0.64	0.52	0.34	0.21
102	College of the Holy Cross	MA	Sep–Oct 2017	0.47	92–144	0.39	0.54	0.49	0.47	0.31	0.45

Rank	Campus	State	Primary window	VCEI	90% rank interval	Dimension scores			Alternative specifications		
						Green	Pedes.	Low car	Arith.	Ver. A	Leaf-on
103	Michigan State University	MI	Nov 2007	0.47	55–169	0.33	0.45	0.69	0.49	0.33	0.47
104	University of Iowa	IA	Sep–Oct 2012	0.47	28–188	0.22	0.86	0.53	0.54	0.34	0.59
105	Virginia Polytechnic Institute and State University	VA	Jun–Jul 2019	0.47	64–172	0.53	0.29	0.67	0.50	0.45	0.47
106	Illinois Institute of Technology	IL	Jul 2018	0.47	74–167	0.59	0.43	0.40	0.47	0.47	0.47
107	University of Michigan-Ann Arbor	MI	Sep–Oct 2014	0.46	88–151	0.38	0.56	0.45	0.47	0.33	0.49
108	Washington College	MD	Oct 2022	0.46	79–167	0.60	0.33	0.50	0.48	0.47	0.35
109	Rensselaer Polytechnic Institute	NY	Sep–Oct 2007	0.46	74–163	0.39	0.60	0.41	0.47	0.29	0.46
110	University of Dayton	OH	Jul–Aug 2007	0.46	68–162	0.34	0.63	0.46	0.48	0.32	0.46
111	Harvard University	MA	Aug–Sep 2013	0.46	56–167	0.34	0.69	0.42	0.48	0.28	0.46
112	University of Missouri-Kansas City	MO	May–Jun 2011	0.46	104–148	0.44	0.45	0.47	0.46	0.30	0.46
113	United States Naval Academy	MD	Aug 2019	0.45	55–174	0.32	0.69	0.42	0.48	0.45	0.45
114	University of Notre Dame	IN	Aug–Sep 2011	0.45	76–178	0.53	0.28	0.63	0.48	0.21	0.45
115	Mount Holyoke College	MA	Oct 2019	0.45	107–154	0.46	0.41	0.49	0.45	0.45	0.16
116	George Mason University	VA	Aug–Sep 2014	0.45	56–193	0.71	0.22	0.57	0.50	0.22	0.45
117	Clark University	MA	Aug 2007	0.44	15–200	0.28	1.00	0.31	0.53	0.33	0.44
118	University of Washington-Bothell Campus	WA	Aug 2011	0.44	98–169	0.52	0.34	0.49	0.45	0.30	0.44
119	Harvey Mudd College	CA	Apr–May 2017	0.44	46–203	0.40	0.73	0.29	0.47	0.28	0.33
120	University of Oregon	OR	Aug–Sep 2011	0.44	97–178	0.43	0.52	0.37	0.44	0.29	0.44
121	Hamilton College	NY	Sep 2013	0.44	17–206	0.71	0.14	0.86	0.57	0.16	0.44
122	Washington & Jefferson College	PA	Apr–May 2012	0.43	67–189	0.24	0.63	0.54	0.47	0.27	0.53
123	Bowdoin College	ME	Jul 2011	0.43	44–198	0.77	0.18	0.59	0.51	0.18	0.43
124	University of California-Berkeley	CA	Mar–Apr 2019	0.43	80–192	0.60	0.24	0.55	0.46	0.43	0.49
125	University of Delaware	DE	May–Jun 2018	0.42	124–170	0.43	0.39	0.45	0.42	0.28	0.42
126	University of Virginia-Main Campus	VA	Oct 2015	0.42	103–177	0.52	0.36	0.41	0.43	0.24	0.46
127	Morehouse College	GA	Aug 2008	0.42	14–207	0.16	1.00	0.46	0.54	0.32	0.42
128	Binghamton University	NY	Aug 2011	0.42	85–195	0.58	0.22	0.58	0.46	0.20	0.42
129	Dickinson College	PA	Apr 2012	0.42	96–183	0.35	0.57	0.37	0.43	0.24	0.30
130	University of California-Los Angeles	CA	Apr–May 2018	0.42	86–189	0.34	0.60	0.35	0.43	0.41	0.36
131	Stony Brook University	NY	Aug 2011	0.41	22–213	0.88	0.12	0.68	0.56	0.17	0.41
132	Hillsdale College	MI	Jul 2014	0.41	42–210	0.74	0.13	0.70	0.52	0.15	0.41
133	Stevens Institute of Technology	NJ	Oct–Nov 2017	0.41	124–182	0.39	0.44	0.40	0.41	0.18	0.37
134	Rutgers University-Camden	NJ	Jun–Jul 2012	0.41	29–204	0.20	0.91	0.36	0.49	0.26	0.41
135	University of California-Riverside	CA	Jun–Jul 2012	0.41	130–182	0.32	0.46	0.45	0.41	0.28	0.41
136	Fordham University	NY	Oct–Nov 2017	0.40	31–217	0.28	0.90	0.26	0.48	0.26	0.49
137	Stanford University	CA	Apr–May 2019	0.40	107–194	0.51	0.24	0.54	0.43	0.40	0.35
138	University of California-Santa Barbara	CA	Oct–Nov 2017	0.40	105–192	0.27	0.43	0.56	0.42	0.29	0.32
139	Scripps College	CA	Aug 2011	0.40	118–190	0.48	0.32	0.41	0.41	0.20	0.40
140	Macalester College	MN	Oct 2016	0.40	123–195	0.40	0.46	0.34	0.40	0.23	0.42

Rank	Campus	State	Primary window	VCEI	90% rank interval	Dimension scores			Alternative specifications		
						Green	Pedes.	Low car	Arith.	Ver. A	Leaf-on
141	University of Minnesota-Crookston	MN	Sep–Oct 2008	0.40	60–200	0.27	0.33	0.71	0.44	0.26	0.26
142	Pomona College	CA	Feb–Mar 2019	0.40	122–192	0.46	0.36	0.38	0.40	0.40	0.45
143	Pitzer College	CA	Apr–May 2018	0.39	99–197	0.32	0.58	0.33	0.41	0.38	0.38
144	University of North Georgia	GA	Jun–Jul 2019	0.39	112–194	0.51	0.27	0.45	0.41	0.20	0.39
145	West Chester University of Pennsylvania	PA	Oct–Nov 2017	0.39	121–189	0.32	0.51	0.37	0.40	0.25	0.59
146	University of Rochester	NY	Jul–Aug 2015	0.39	141–190	0.39	0.35	0.43	0.39	0.22	0.39
147	California Institute of Technology	CA	Sep–Oct 2011	0.39	140–192	0.39	0.33	0.46	0.39	0.19	0.31
148	Occidental College	CA	Sep 2007	0.39	133–193	0.28	0.43	0.48	0.40	0.24	0.39
149	Lycoming College	PA	Aug–Sep 2018	0.39	95–204	0.32	0.59	0.31	0.41	0.20	0.39
150	Georgetown University	DC	Jul 2013	0.38	112–201	0.22	0.49	0.52	0.41	0.24	0.38
151	Florida State University	FL	Nov 2007	0.38	40–209	0.17	0.82	0.39	0.46	0.28	0.47
152	Southwestern University	TX	May 2022	0.38	60–212	0.71	0.13	0.58	0.47	0.40	0.38
153	Texas A & M University-College Station	TX	Nov–Dec 2007	0.38	45–214	0.16	0.78	0.44	0.46	0.31	0.53
154	Villanova University	PA	Jul 2019	0.38	121–199	0.49	0.26	0.43	0.39	0.36	0.38
155	Purdue University-Main Campus	IN	Sep 2007	0.38	48–211	0.17	0.76	0.42	0.45	0.31	0.38
156	University of California-Irvine	CA	Jun–Jul 2011	0.38	139–197	0.35	0.32	0.49	0.39	0.24	0.38
157	University of California-Davis	CA	Apr–May 2019	0.38	131–202	0.46	0.32	0.36	0.38	0.38	0.46
158	University of North Carolina at Charlotte	NC	Oct 2012	0.37	47–217	0.74	0.10	0.67	0.51	0.16	0.46
159	University of Denver	CO	Jul–Aug 2018	0.37	145–201	0.35	0.42	0.35	0.37	0.20	0.37
160	Northwestern University	IL	Oct–Nov 2017	0.37	156–200	0.32	0.37	0.42	0.37	0.20	0.38
161	Franklin and Marshall College	PA	Jun 2011	0.37	83–211	0.64	0.21	0.36	0.40	0.16	0.37
162	University of Houston	TX	Oct–Nov 2018	0.36	124–201	0.23	0.53	0.40	0.39	0.36	0.40
163	University of Minnesota-Twin Cities	MN	Sep–Oct 2009	0.36	52–215	0.23	0.76	0.28	0.42	0.26	0.39
164	San Diego State University	CA	Jul–Aug 2018	0.36	95–207	0.19	0.61	0.41	0.40	0.24	0.36
165	Auburn University	AL	Jun–Jul 2013	0.36	85–213	0.60	0.14	0.58	0.44	0.15	0.36
166	Northeastern University	MA	Jul–Aug 2017	0.36	77–211	0.18	0.65	0.42	0.41	0.21	0.36
167	Central Community College	NE	Jul–Aug 2019	0.36	115–210	0.24	0.35	0.57	0.39	0.41	0.36
168	Berea College	KY	Jun 2013	0.36	86–210	0.62	0.15	0.50	0.42	0.13	0.36
169	University of Pittsburgh-Pittsburgh Campus	PA	Jun–Jul 2011	0.36	55–223	0.25	0.75	0.24	0.41	0.18	0.36
170	California State University-Fullerton	CA	Jul–Aug 2015	0.36	89–212	0.24	0.64	0.29	0.39	0.26	0.36
171	Carnegie Mellon University	PA	Oct–Nov 2017	0.35	17–226	0.10	1.00	0.43	0.51	0.23	0.40
172	Syracuse University	NY	Oct–Nov 2020	0.35	44–223	0.12	0.79	0.46	0.45	0.35	0.40
173	Clarkson University	NY	Aug–Sep 2018	0.34	84–217	0.61	0.11	0.57	0.43	0.14	0.34
174	Georgia Institute of Technology-Main Campus	GA	Oct–Nov 2007	0.34	48–218	0.16	0.80	0.30	0.42	0.24	0.42
175	Massachusetts Institute of Technology	MA	Sep–Oct 2013	0.34	51–219	0.15	0.78	0.33	0.42	0.22	0.37
176	State University of New York at New Paltz	NY	Nov 2015	0.34	160–212	0.32	0.27	0.45	0.35	0.21	0.38
177	Johns Hopkins University	MD	Aug–Sep 2018	0.34	100–216	0.58	0.15	0.43	0.39	0.11	0.34
178	Loyola University Chicago	IL	Sep–Oct 2015	0.34	95–221	0.14	0.60	0.45	0.40	0.19	0.38

Rank	Campus	State	Primary window	VCEI	90% rank interval	Dimension scores			Alternative specifications		
						Green	Pedes.	Low car	Arith.	Ver. A	Leaf-on
179	Colorado College	CO	Sep–Oct 2014	0.33	163–217	0.37	0.30	0.32	0.33	0.17	0.24
180	Stonehill College	MA	Aug 2021	0.33	21–223	0.86	0.05	0.77	0.56	0.33	0.33
181	University of Maryland-Baltimore County	MD	Jun–Jul 2019	0.33	121–221	0.51	0.19	0.36	0.35	0.30	0.33
182	Virginia Commonwealth University	VA	May–Jun 2017	0.32	69–222	0.17	0.72	0.28	0.39	0.20	0.32
183	Temple University	PA	Aug–Sep 2014	0.32	22–231	0.10	1.00	0.33	0.48	0.09	0.32
184	Claremont McKenna College	CA	Feb–Mar 2019	0.31	142–225	0.23	0.50	0.26	0.33	0.31	0.39
185	Hope College	MI	Aug–Sep 2012	0.31	151–226	0.43	0.18	0.38	0.33	0.10	0.31
186	University of California-San Diego	CA	Feb–Mar 2020	0.30	168–221	0.24	0.39	0.30	0.31	0.30	0.36
187	University of Central Oklahoma	OK	Dec 2007–Jan 2008	0.30	22–235	0.08	1.00	0.34	0.47	0.24	0.35
188	University of Chicago	IL	Jul–Aug 2018	0.30	161–232	0.38	0.24	0.23	0.30	0.29	0.30
189	University of Massachusetts-Amherst	MA	Oct 2018	0.30	82–223	0.58	0.07	0.62	0.42	0.14	0.26
190	Colorado School of Mines	CO	Aug 2008	0.30	183–223	0.24	0.33	0.33	0.30	0.17	0.30
191	New York University	NY	Jun 2019	0.29	62–232	0.10	0.74	0.35	0.40	0.29	0.29
192	Pepperdine University	CA	Apr 2019	0.29	188–228	0.27	0.26	0.35	0.30	0.30	0.16
193	University of Vermont	VT	Jul–Aug 2011	0.29	148–230	0.43	0.14	0.41	0.33	0.10	0.29
194	George Washington University	DC	Jul 2009	0.29	147–227	0.15	0.47	0.36	0.32	0.15	0.29
195	Bucknell University	PA	Aug 2018	0.29	71–224	0.68	0.06	0.57	0.44	0.14	0.29
196	Vanderbilt University	TN	Mar–Apr 2014	0.29	172–227	0.19	0.39	0.32	0.30	0.16	0.42
197	Loyola Marymount University	CA	Oct 2011	0.28	189–231	0.29	0.25	0.31	0.28	0.13	0.36
198	Howard University	DC	Jul–Aug 2009	0.28	68–231	0.11	0.74	0.26	0.37	0.16	0.28
199	California State University-Long Beach	CA	Dec 2017–Jan 2018	0.28	182–233	0.18	0.34	0.36	0.29	0.28	0.30
200	University of San Diego	CA	Mar–Apr 2009	0.27	193–233	0.25	0.22	0.37	0.28	0.11	0.15
201	University of Central Florida	FL	Nov 2011	0.27	135–234	0.47	0.11	0.39	0.32	0.12	0.33
202	Juniata College	PA	Aug 2018	0.27	84–228	0.51	0.06	0.65	0.40	0.13	0.27
203	The University of Tennessee-Knoxville	TN	Dec 2017–Jan 2018	0.26	69–237	0.09	0.74	0.28	0.37	0.18	0.31
204	Willamette University	OR	Aug 2011	0.26	138–237	0.50	0.12	0.30	0.31	0.11	0.26
205	Gustavus Adolphus College	MN	Oct 2015	0.26	93–232	0.41	0.06	0.66	0.38	0.12	0.39
206	University of Southern California	CA	Feb–Mar 2015	0.26	92–237	0.12	0.68	0.21	0.34	0.15	0.24
207	University of Kentucky	KY	Nov 2017	0.26	180–237	0.17	0.38	0.26	0.27	0.17	0.52
208	Marquette University	WI	Sep–Oct 2019	0.25	59–238	0.08	0.79	0.25	0.37	0.26	0.24
209	University of the Pacific	CA	Jan–Feb 2009	0.25	175–237	0.16	0.40	0.24	0.27	0.19	0.39
210	The University of Texas at Austin	TX	Oct–Nov 2007	0.25	163–240	0.16	0.46	0.22	0.28	0.18	0.27
211	University of Illinois Chicago	IL	Apr–May 2009	0.24	41–242	0.08	0.90	0.19	0.39	0.14	0.33
212	New Jersey Institute of Technology	NJ	Oct 2009	0.24	135–239	0.09	0.55	0.27	0.30	0.13	0.19
213	Gonzaga University	WA	Jul–Aug 2011	0.24	199–241	0.27	0.19	0.25	0.24	0.10	0.24
214	Washington University in St Louis	MO	Aug 2018	0.23	181–240	0.13	0.39	0.26	0.26	0.13	0.23
215	Hobart William Smith Colleges	NY	Aug–Sep 2013	0.23	41–231	0.77	0.02	0.68	0.49	0.16	0.23
216	Rutgers University-Newark	NJ	Jun–Jul 2018	0.23	80–245	0.10	0.73	0.17	0.33	0.23	0.23
217	Yeshiva University	NY	Aug–Sep 2014	0.23	155–244	0.13	0.49	0.19	0.27	0.06	0.23
218	Boston University	MA	Jul–Aug 2009	0.23	183–242	0.11	0.37	0.30	0.26	0.11	0.23

Rank	Campus	State	Primary window	VCEI	90% rank interval	Dimension scores			Alternative specifications		
						Green	Pedes.	Low car	Arith.	Ver. A	Leaf-on
219	Indiana University-Purdue University-Indianapolis	IN	May–Jun 2019	0.23	198–242	0.16	0.31	0.24	0.23	0.22	0.23
220	University of San Francisco	CA	Oct–Nov 2017	0.22	201–246	0.20	0.28	0.19	0.22	0.23	0.34
221	Bard College	NY	May 2012	0.22	2–235	1.00	0.00	1.00	0.67	0.19	0.22
222	Bryn Mawr College	PA	May–Jun 2022	0.21	23–234	0.86	0.01	0.76	0.54	0.19	0.21
223	University of Pennsylvania	PA	Oct–Nov 2016	0.21	109–243	0.06	0.66	0.21	0.31	0.11	0.24
224	United States Military Academy	NY	May 2023	0.21	5–236	1.00	0.00	0.83	0.63	0.21	0.21
225	Middlebury College	VT	Apr 2012	0.20	141–239	0.33	0.05	0.55	0.31	0.11	0.45
226	University of South Florida	FL	Apr–May 2012	0.20	163–245	0.45	0.08	0.22	0.25	0.10	0.27
227	Williams College	MA	Aug–Sep 2018	0.20	50–233	0.63	0.02	0.75	0.47	0.15	0.20
228	Creighton University	NE	Jun–Jul 2019	0.20	111–244	0.05	0.66	0.24	0.31	0.19	0.20
229	Saint Louis University	MO	Feb–Mar 2022	0.19	101–248	0.07	0.70	0.16	0.31	0.20	0.21
230	Lewis & Clark College	OR	Aug–Sep 2011	0.19	12–239	0.95	0.00	0.76	0.57	0.16	0.19
231	Furman University	SC	Jul–Aug 2012	0.19	15–239	0.88	0.00	0.81	0.56	0.18	0.19
232	University of California-Santa Cruz	CA	May–Jun 2011	0.19	15–239	0.91	0.00	0.79	0.57	0.18	0.19
233	Vassar College	NY	Jul 2018	0.19	15–240	0.93	0.00	0.72	0.55	0.17	0.19
234	St. Mary's College of Maryland	MD	May 2018	0.19	18–241	0.73	0.00	0.88	0.54	0.17	0.19
235	Florida International University	FL	May 2012	0.18	168–242	0.35	0.04	0.44	0.28	0.11	0.18
236	Drexel University	PA	Oct–Nov 2018	0.17	58–247	0.03	0.80	0.24	0.36	0.17	0.25
237	University of New Hampshire-Main Campus	NH	Aug 2015	0.17	38–242	0.82	0.00	0.62	0.48	0.16	0.17
238	University of Massachusetts-Boston	MA	Jun–Jul 2019	0.16	191–245	0.12	0.07	0.45	0.21	0.17	0.16
239	Rochester Institute of Technology	NY	Oct 2010	0.16	63–244	0.53	0.00	0.72	0.42	0.15	0.61
240	Saint Mary's College of California	CA	Feb–Mar 2020	0.15	94–244	0.62	0.00	0.51	0.37	0.15	0.51
241	University of Maryland, Baltimore	MD	Oct–Nov 2017	0.14	135–248	0.02	0.60	0.26	0.29	0.09	0.21
242	Virginia Military Institute	VA	Oct 2015	0.13	137–245	0.45	0.00	0.49	0.31	0.12	0.13
243	Soka University of America	CA	Apr 2019	0.13	148–246	0.43	0.00	0.48	0.30	0.13	—
244	University of Puget Sound	WA	Oct 2012	0.11	177–247	0.41	0.00	0.29	0.23	0.08	0.60
245	Lane Community College	OR	Aug–Sep 2011	0.10	199–248	0.30	0.00	0.29	0.20	0.10	0.10
246	Barnard College	NY	Sep 2015	0.09	232–249	0.12	0.05	0.14	0.10	0.02	0.09
247	Columbia University in the City of New York	NY	Apr–May 2009	0.09	228–249	0.04	0.08	0.24	0.12	0.03	0.07
248	University of California-Merced	CA	Apr–May 2019	0.09	218–249	0.00	0.22	0.28	0.17	0.09	0.10
249	United States Air Force Academy	CO	Jul–Aug 2015	0.05	222–249	0.04	0.00	0.39	0.14	0.05	0.05

Version B, primary window; the 249 campuses with a usable primary window, ordered by rank (1 = highest index). VCEI: geometric mean of the three dimension scores (floor 0.01), the index reported in Section 5.7. 90% rank interval: 5th to 95th percentile of the campus's rank over 500 draws that vary the dimension weights, drop one indicator, and switch the aggregation rule. Dimension scores, each from 0 to 1 and the equal-weight mean of its indicators normalized with the goalposts of Table S4: Green, greenery; Pedes., pedestrian space; Low car, low car dominance (the car-dominance indicators are reversed, so higher values mean less car dominance). Alternative specifications: Arith., the arithmetic instead of the geometric mean of the dimension scores; Ver. A, the geometric index from all panoramas in the primary window with stored road and sidewalk shares (Version A); Leaf-on, the geometric index (Version B) in the campus's best-covered window entirely within May–September (—: no such window). Primary window: the months of the window with the highest coverage. Colorado Mountain College, Principia College, and the University of Minnesota-Rochester have no usable primary window and no index value.

Figures S1–S8

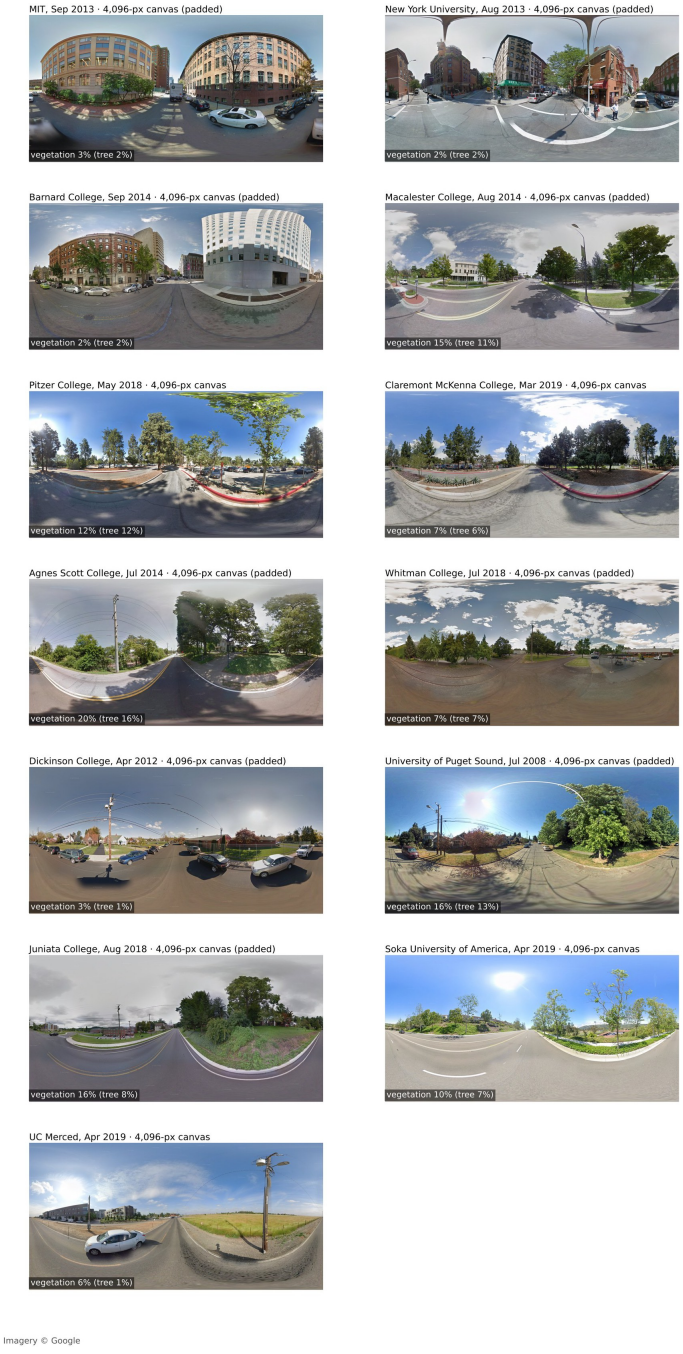

Figure S1: All 13 panoramas of the random gallery. Full 360° views (seed 20261028) with padding removed and the vegetation shares used in the analysis; none contains a special-object detection at confidence $\geqslant$ 0.80.

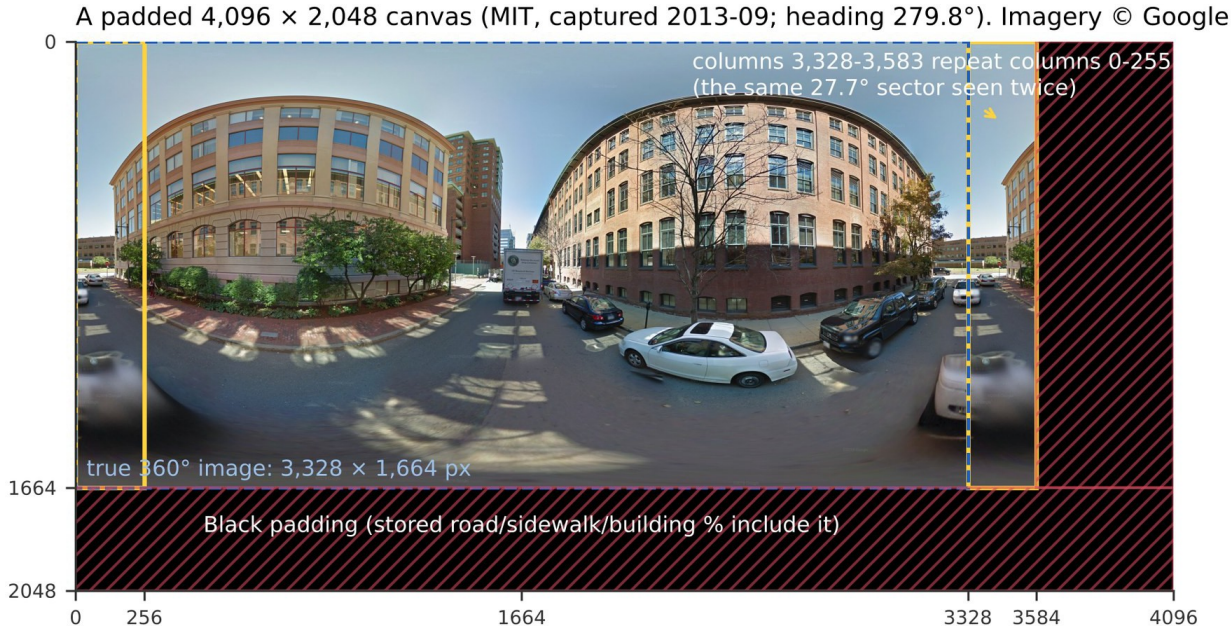

Figure S2: A padded 4,096 × 2,048 canvas. The true panorama is 3,328 × 1,664 px; columns 3,328–3,583 repeat columns 0–255, so a 27.7° sector appears twice; the rest is black.

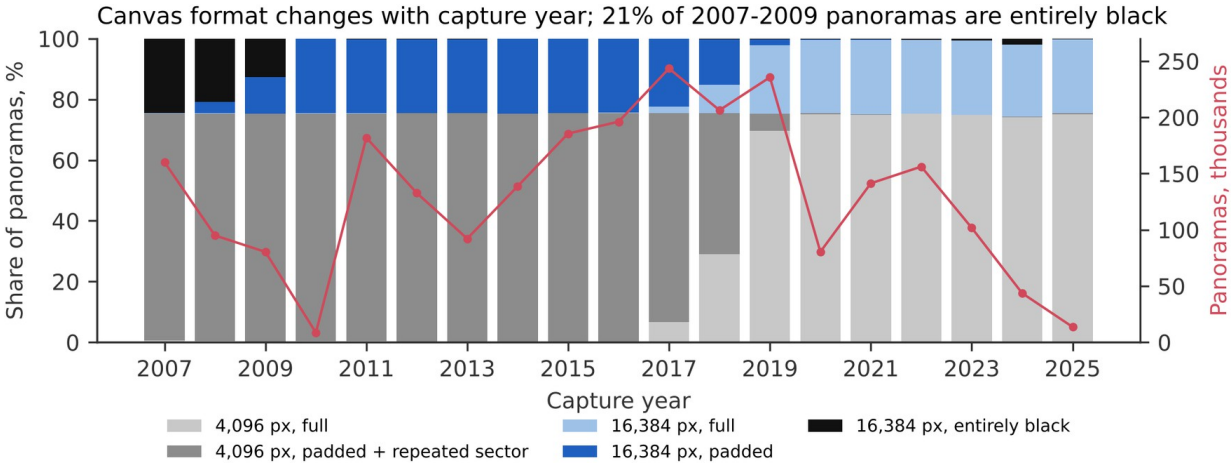

Figure S3: Canvas format by capture year. All 2,494,281 files. Padded canvases dominate captures through 2018 and full canvases from 2019 on; about a quarter of each year's files are 16,384 px wide, and most entirely black files are 2007–2009 captures.

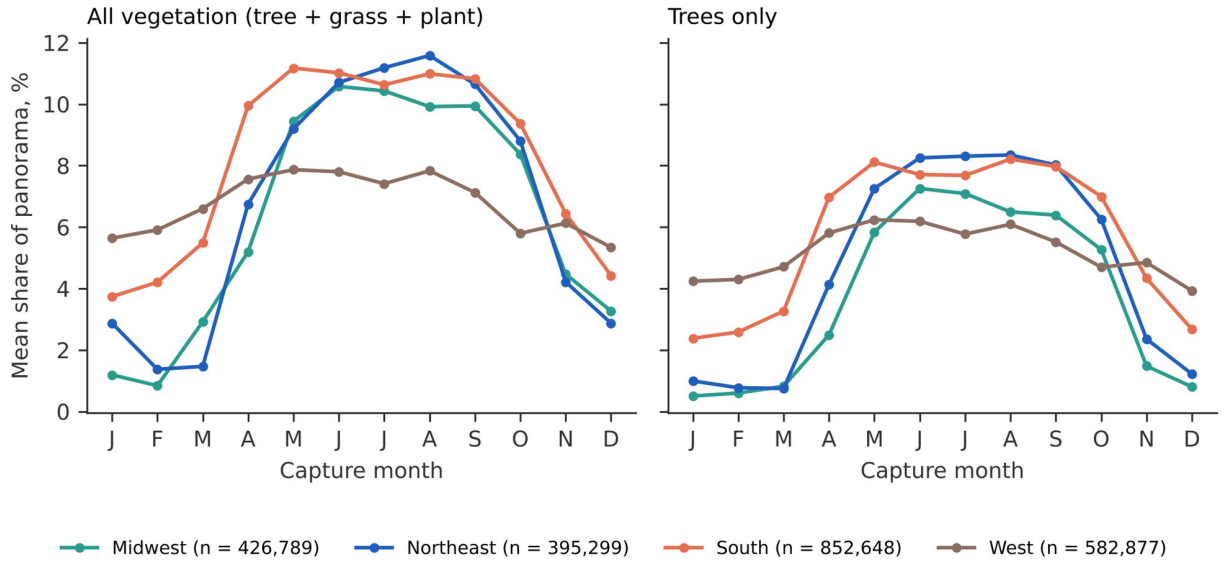

Figure S4: Vegetation share by capture month and census region. Mean over all 2.26 million panoramas in the corpus with a vegetation measure. The seasonal swing is largest in the Midwest and Northeast and smallest in the West.

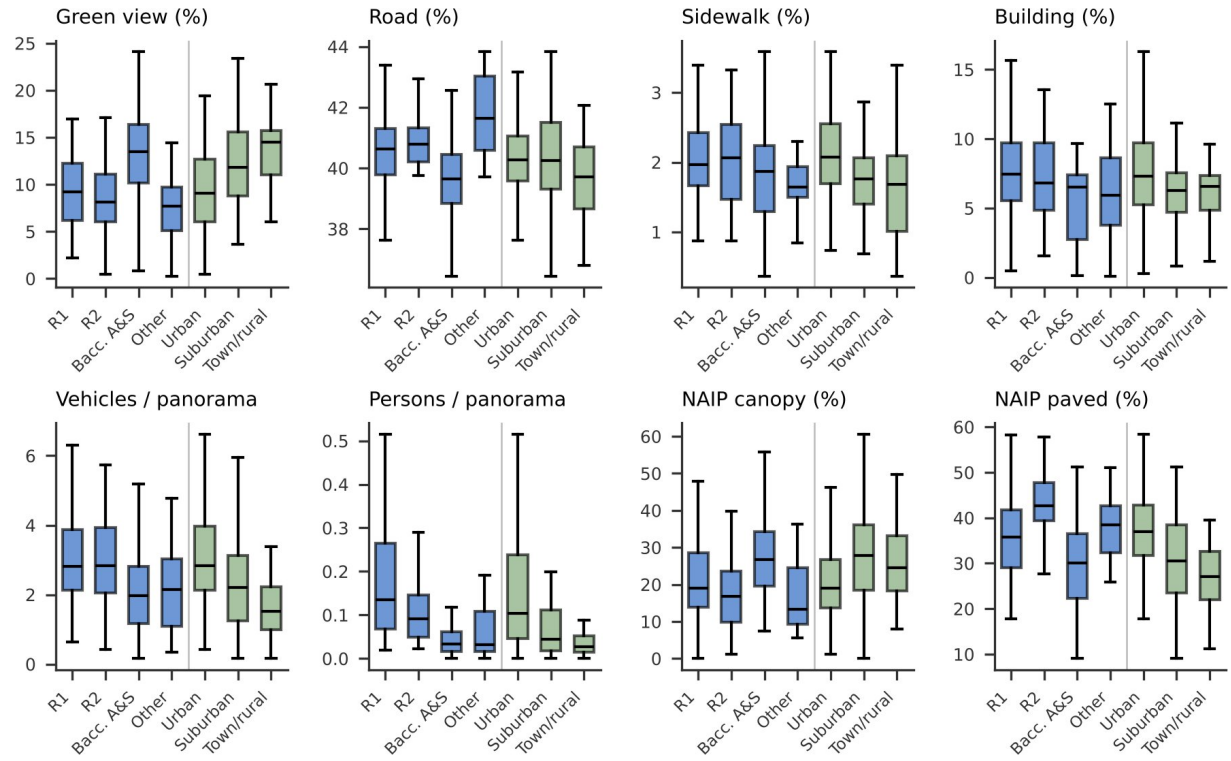

Figure S5: Campus-level measures by Carnegie class and locale. Version B, primary window; Carnegie class in blue, locale in green. Boxes show the interquartile range and median; whiskers extend to 1.5 IQR.

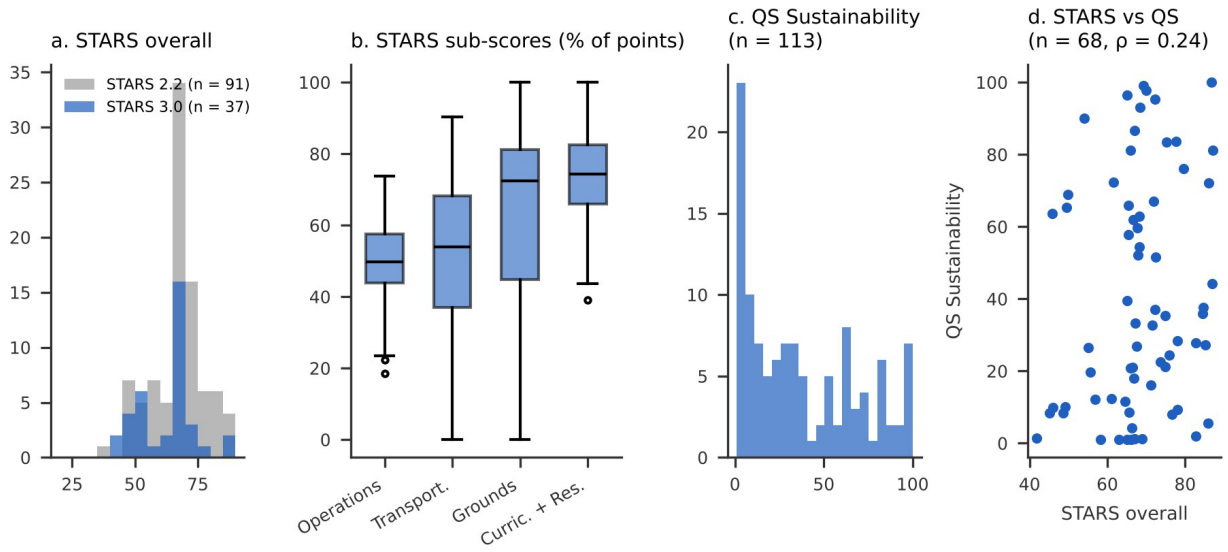

Figure S6: Outcome distributions. (a) STARS overall score by STARS version. (b) STARS sub-scores used in the analysis. (c) QS Sustainability. (d) STARS overall against QS Sustainability for campuses with both.