Brian DeVoe GIS - EE505

**Developing a Longitudinal Census-Tract Greenness Indicator for COI**

**(NOTE: ChatGPT was used to help construct this project outline but not the general idea, goal, methods, purpose, etc.)**

**1. Purpose and Motivation**

Existing child opportunity and environmental indices often rely on cross-sectional or city-level measures of green space (e.g., NatureScore). They lack a **consistent, longitudinal measure of greenness at the census-tract scale**, which limits understanding of how changing vegetation and park access affect health and opportunity over time.

Creating a **yearly, tract-level greenness time series** will allow:

* Detection of spatial and temporal trends in neighborhood vegetation (e.g., canopy growth or loss, post-disaster recovery).
* Integration with annual health and socioeconomic indicators (e.g., life expectancy, obesity, mental health).
* Use in modeling frameworks that exploit within-tract change (e.g., fixed-effects, lagged exposures).

**2. Data Sources**

| **Indicator** | **Dataset / Source** | **Notes** |
| --- | --- | --- |
| **NDVI & EVI** | Landsat Surface Reflectance, USGS ([usgs.gov](https://www.usgs.gov/landsat-missions/landsat-enhanced-vegetation-index)) | Annual/seasonal composites (30 m); cloud-masked and harmonized via Earth Engine |
| **Tree Canopy Cover (TCC)** | USDA Forest Service Tree Canopy Cover ([data.fs.usda.gov](https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/)) | 30 m gridded canopy fraction (1985–2023) |
| **Land Cover Classes** | USGS National Land Cover Database ([mrlc.gov](https://www.mrlc.gov/)) | Used to compute % vegetated area (forest, grass, shrub, wetland) |
| **Park Proximity / Access** | PAD-US-AR (Browning et al., 2022), ParkServe, CDC-derived tract park data ([icpsr.umich.edu](https://www.icpsr.umich.edu/sites/nanda/view/studies/117921)) | Percent of tract (and 0.5-mile buffer) considered park area |
| **Geographies** | Census tracts, block groups, and counties — U.S. Census Bureau TIGER/Line shapefiles | Consistent boundary version (2020) across years |

*Potential additions:* Sentinel-2 indices (10 m), MODIS Vegetation Continuous Fields (250 m), Global Forest Change, or population-weighted greenness exposures (LandScan / WorldPop)

**3. Data Harmonization and Aggregation Process**

**Goal**

Create a uniform, yearly “greenness stack” — all indicators aligned to the same spatial grid — and compute tract-level summaries via zonal statistics.

**Why raster harmonization**

* All major inputs (NDVI, EVI, TCC, land cover) are *raster-based*.
* Aligning them to one 30 m Albers (EPSG:5070) grid ensures that every pixel represents the same geographic footprint across datasets and years.
* This prevents spatial mismatch and allows reproducible tract, block-group, or county aggregation.

**Process Overview**

1. **Define a master grid**
   * CRS: EPSG 5070 (Albers Equal-Area)
   * Resolution: 30 m
   * Alignment: match to base NDVI mosaic
2. **Resample each input raster**
   * Continuous layers (NDVI, EVI, TCC): *bilinear* resampling
   * Categorical layers (land cover): *nearest-neighbor*
   * Park polygons: rasterized to fractional % coverage per 30 m pixel
3. **Stack and harmonize**
   * Build annual raster stacks (or VRTs) with identical extent, resolution, and nodata masks.
   * Optionally store as cloud-optimized GeoTIFFs.
4. **Zonal aggregation**
   * Overlay census tract polygons (and other geographies).
   * Compute mean NDVI, mean EVI, mean TCC, % vegetated area, % park area (and 0.5-mile buffer variant).
   * Output one record per tract per year, including coverage diagnostics.

**Outcome**

A tract-year panel dataset (≈ 70 k tracts × 20 years = 1.4 M records) providing consistent greenness metrics across time, ready for index construction and validation.

**4. Challenges in Harmonization**

* **Temporal mismatch:** different update frequencies (Landsat yearly, NLCD every 2–3 years).
* **Boundary drift:** census tract definitions evolve (2010 → 2020); will require harmonized boundary crosswalk.
* **Resolution conflicts:** reprojecting 10 m–250 m sources to a common 30 m grid.
* **Missing or cloud-masked pixels:** need minimum valid-pixel thresholds or interpolation.
* **Rural extremes:** extremely high greenness may dominate scales; apply top-coding?
* **Processing scale:** national mosaics require efficient parallel / tiled processing (e.g., VRT + windowed reads).

**5. Methods and Weighting Strategy**

**Index Construction**

* **Equal-weight baseline:** each standardized indicator contributes equally.
* **Expert-weight model:** emphasize tree canopy and park proximity.
* **PCA-weighted model:** use unsupervised PCA loadings (PC1) to empirically derive weights and reduce collinearity.

**Validation**

Use tract-level health and well-being data:

* CDC PLACES (obesity, mental health, physical activity)
* NCHS life expectancy
* Asthma / air-quality data (EPA AQS, PLACES)
* Education outcomes (SEDA test scores)

Evaluate models via correlations, regression R², predictive validity, and spatial residual diagnostics.  
Select the index version that balances interpretability and predictive performance.

**6. Expected Deliverables**

1. Annual tract-level CSV + GPKG files (2000–present) \*\*For this class 1 year of data\*\*
2. Technical documentation describing harmonization and weighting
3. Validation tables / figures linking greenness to health outcomes
4. Reproducible scripts (Python + Earth Engine) for future updates