surveyzones: Workload-Balanced Geographically Bounded Zone Construction for Survey Field Operations

Eduardo Leoni, IBGE

February 2026

Introduction

Survey organizations face a fundamental operational challenge: how to partition geographic regions into zones for field teams such that each team has a manageable, balanced workload; field operations remain efficient (teams don't waste time traveling); and zones are geographically cohesive and traversable within a reasonable time budget.

Existing facility location models in the operations research literature—such as the Location Set Covering Problem (LSCP), the p-median problem, and the Warehouse Location Problem (WLP)—were designed for retail, emergency services, or supply chain contexts. No single model jointly addresses the three constraints specific to survey field operations: workload balance, a hard geographic diameter bound, and automatic zone-count discovery.

Key finding. surveyzones is the only model that jointly enforces workload capacity, a hard geographic diameter bound ($D_{\max}$), and automatic minimum-K discovery. Each existing model satisfies at most one of these three constraints.

Survey-Specific Constraints

Workload balance. Field teams have finite capacity. Zones must respect a maximum workload per team (`max_workload_per_zone`). Unlike logistics applications where per-stop demand is unknown in advance, survey workload is largely predetermined: the target number of households per tract is fixed by the sampling frame. Expected service time per tract is therefore estimable before deployment, making workload-balanced partitioning feasible.

Geographic reach. Teams cannot efficiently operate in zones that are geographically dispersed. $D_{\max}$ defines the maximum travel time allowed between any tract and its zone center. Any tract more than $D_{\max}$ minutes from the zone center is excluded from that zone. This implicitly bounds zone diameter: the maximum inter-tract travel time within a zone is at most $2 \times D_{\max}$.

Related Work

p-Median

The p-median problem minimizes total weighted travel distance from demand points to their nearest open facility, given K facilities:

$$\min \sum_i \sum_j d_{ij} \cdot w_i \cdot x_{ij}$$

subject to $\sum_j y_j = K$ and $\sum_j x_{ij} = 1 \ \forall i$.

p-Median minimizes total travel distance (demand-weighted by w_i); surveyzones minimizes unweighted total travel time, with demand entering only through the workload capacity constraint. Both minimize a sum-of-distances objective, but p-median imposes no workload capacity and no zone-size bound. It may produce zones with highly unequal workloads, since geographic proximity—not workload—drives assignment. It also allows zones to span arbitrarily large areas, which is operationally infeasible for survey teams.

Warehouse Location Problem

The Warehouse Location Problem (WLP) minimizes fixed facility-opening costs plus variable transportation costs, subject to warehouse capacity:

$$\min \sum_j f_j y_j + \sum_i \sum_j c_{ij} x_{ij}$$

subject to $\sum_j x_{ij} = 1 \ \forall i$ and $\sum_i s_i x_{ij} \leq S_j y_j \ \forall j$.

surveyzones is structurally a capacitated facility location problem: every tract is a candidate zone center, opening costs are uniform, and the assignment cost is travel time. Unlike WLP—where K is determined by trading off fixed opening costs against transport savings—surveyzones finds the minimum K that yields a feasible assignment via an external search loop. The critical addition absent from standard WLP is the $D_{\max}$ constraint: WLP respects capacity but permits zones of arbitrary geographic extent.

Location Set Covering Problem

The Location Set Covering Problem (LSCP) minimizes the number of facilities needed to ensure every demand point is within $D_{\max}$ of some open facility:

$$\min \sum_j y_j \quad \text{s.t.} \quad \sum_{j:\, d_{ij} \leq D_{\max}} y_j \geq 1 \ \forall i.$$

LSCP shares surveyzones' goal of using as few zones as possible, and it uses $D_{\max}$ as a parameter. However, it produces no zone assignments—only facility locations—and its objective contains no workload balancing or distance minimization. Any two configurations achieving the same minimum K are equally optimal, even if one requires twice the total travel time.

Summary

	LSCP	p-median	WLP	**surveyzones**
Workload capacity				
Maximum diameter				
100% assignment	†			
Distance minimization				
Automatic K discovery				

†LSCP ensures coverage (reachability within $D_{\max}$); it does not assign tracts to zones.

The surveyzones Model

Formulation

For a fixed number of zones K, surveyzones solves a capacitated p-median MILP:

$$\min \sum_i \sum_k d_{ik} \cdot z_{ik}$$

subject to:

$$\sum_k z_{ik} = 1 \quad \forall i \quad \text{(all tracts assigned)}$$

$$\sum_k y_k = K \quad \text{(exactly } K \text{ zones)}$$

$$z_{ik} \leq y_k \quad \forall i, k \quad \text{(assign only to open zones)}$$

$$\sum_i s_i \cdot z_{ik} \leq W_{\max} \cdot y_k \quad \forall k \quad \text{(workload capacity)}$$

The geographic reach constraint ($d_{ik} \leq D_{\max}$) is enforced in pre-processing: any (i, k) pair with travel time exceeding $D_{\max}$ is excluded from the model, making z_{ik} infeasible by construction.

K is discovered via a search loop: starting from a lower bound $K_{\min}$ derived from workload and target zone size, K is incremented until the MILP is feasible:

$$K^* = \min\{K \geq K_{\min} : \text{MILP feasible}\}$$

This architecture combines minimum-K search (inspired by LSCP), p-median's distance minimization, WLP's capacity constraints, and a novel $D_{\max}$ pre-filtering step that keeps zones geographically compact without adding MILP constraints.

Implementation

surveyzones operates in four stages.

Stage 1 — Distance computation. Pairwise travel times are computed using OSRM as the primary engine. Where OSRM cannot route a pair (e.g., islands, river-only access), haversine distance is used and scaled to a large time penalty, discouraging cross-water assignments while keeping every tract assignable as a last resort. All distances are stored in a sparse format.

Stage 2 — Connectivity analysis. Distances are filtered by $D_{\max}$ to form a connectivity graph. Disconnected components are identified and partition constraints (e.g., by state) applied via `enforce_partition` are resolved. Each component is passed to Stage 3 as an independent sub-problem.

Stage 3 — Minimum-K search. The lower bound $K_{\min} = \max(\lceil n/\text{target\_zone\_size} \rceil,\ \lceil W_{\text{total}}/W_{\max} \rceil)$ is computed. For $K = K_{\min}, K_{\min} + 1, \ldots$, the capacitated p-median MILP is solved with HiGHS (time limit `max_time`, gap tolerance `rel_tol`). The search returns the first K for which the solver finds a solution within `rel_tol` of optimality.

Stage 4 — Sequencing. Tracts within each zone are ordered by solving an asymmetric TSP, respecting road-network asymmetry. The zones themselves are then sequenced by solving a second ATSP (or nearest-neighbor approximation) over the K zone centers, producing an optimal deployment order for supervisors or coordinators. Zone statistics (diameter, workload) are computed for diagnostic output.

Conclusion

surveyzones fills a gap in the facility location literature. No existing model jointly enforces the three constraints required for survey field operations: workload capacity, a hard geographic diameter bound, and automatic minimum-K discovery. p-Median minimizes total distance but ignores all three constraints. WLP adds workload capacity but lacks both diameter bounds and automatic K discovery. LSCP discovers minimum K but ignores workload and distance. surveyzones extends the capacitated facility location framework with $D_{\max}$ pre-filtering, yielding the only model that satisfies all three constraints simultaneously.

References

- **ompr** — R package for modelling mixed integer linear programs; the warehouse location example in ompr's documentation served as the structural inspiration for surveyzones' MILP formulation. https://dirkschumacher.github.io/ompr/
- **spopt** (PySAL) — Python facility location library covering p-median, p-center, LSCP, MCLP, and capacitated variants. The closest existing tool to surveyzones; lacks hard $D_{\max}$ pre-filtering and automatic K search. https://pysal.org/spopt/
- **OSRM** — Open Source Routing Machine, used for road-network travel time computation. https://project-osrm.org/
- **HiGHS** — Open-source MILP solver used by surveyzones. https://highs.dev/