Esri International User Conference
San Diego, California

Demo Theater | July 24, 2012

Apportioning population between areas

Linda Beale PhD

Apportioning population between areas

Session outline

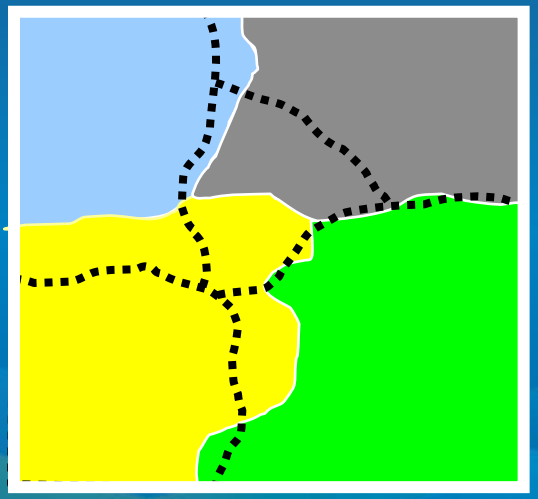
- Population data is usually known at census boundaries
- So how can we shift or apportion data to other boundaries?
  - Outline three different methods
  - Demonstrate ArcGIS workflows
  - Show automated processes
  - Discuss modifiable areal unit problem

Areal interpolation

- Moving area data between different boundaries requires some form of areal interpolation
- Numerous different approaches
  - Solution determined by method and data

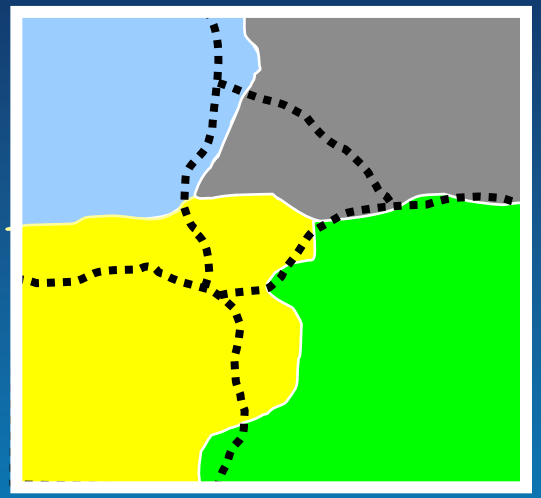
Data to larger areas

- If we are aggregating data to a lower resolution (i.e. larger areas)
  - Geographic boundaries nest within each other
  - We can use Dissolve or Summary Statistics
  - Possibly use spatial join?
- But what if it's a completely different geography?



Methods

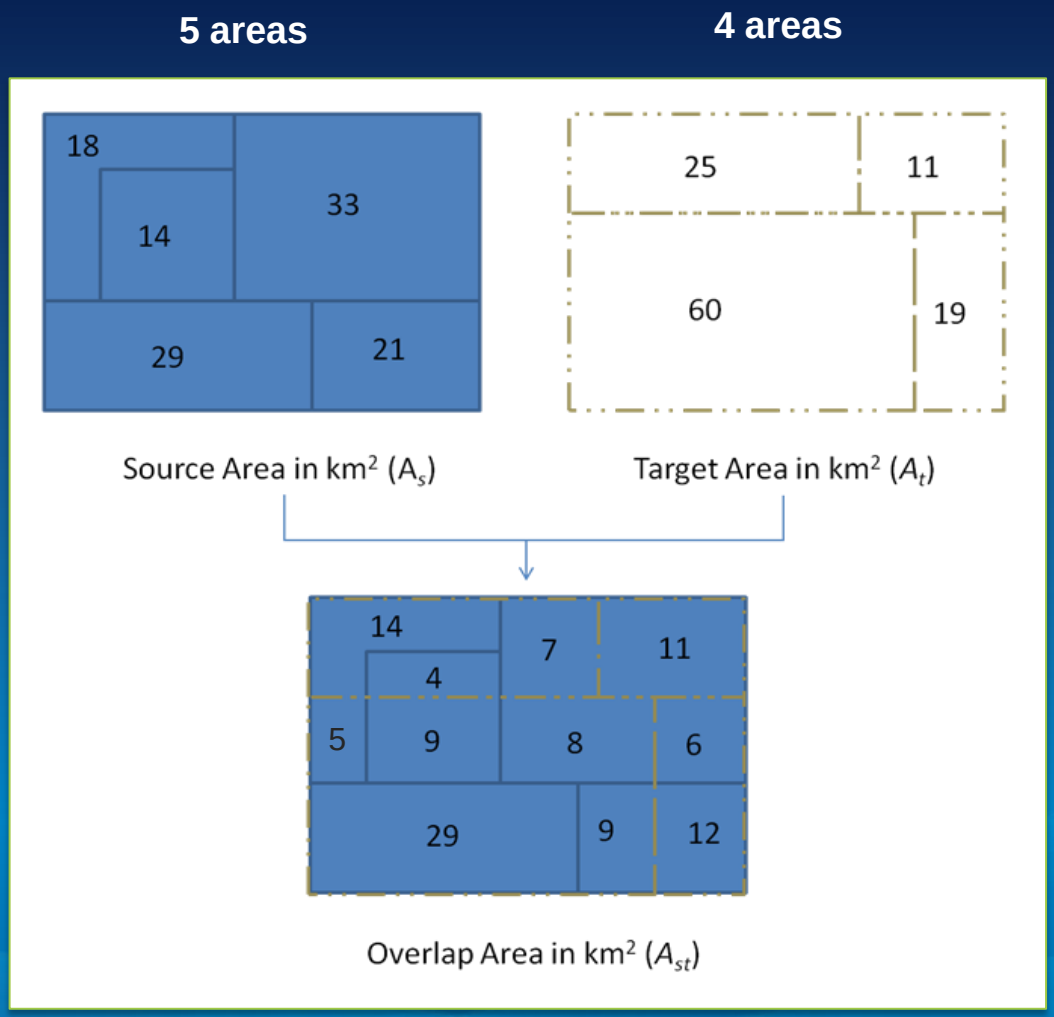
- We will discuss three methods:
  - Simple area weighting
  - Dasymetric approaches
    - Mask area weighting
    - Filtered area weighting



- Data is transferred between:
  - Source zones
  - Target zones
  - Some methods use ancillary data

- Observe the pycnophylactic principle

Simple Area Weighting

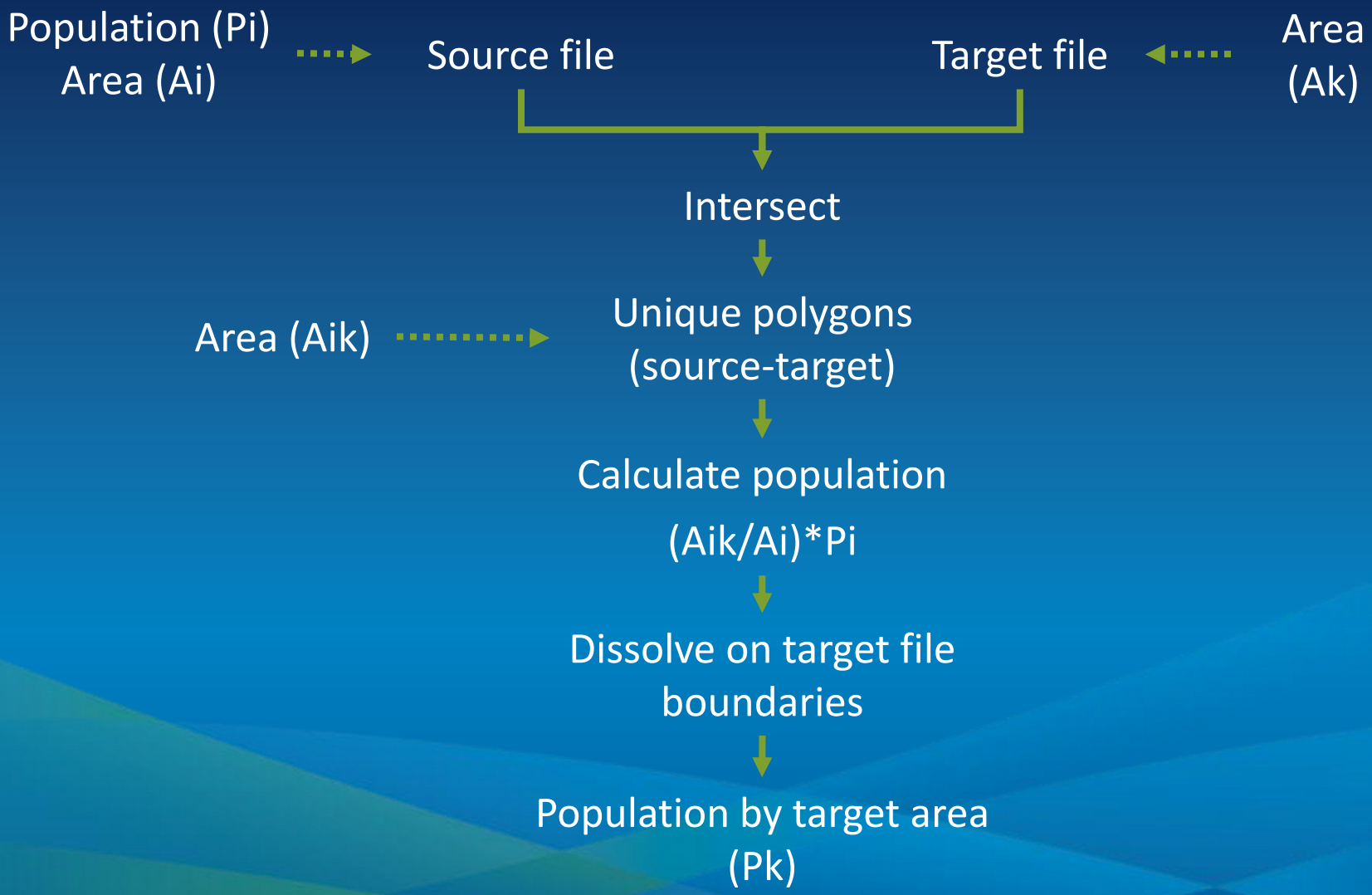


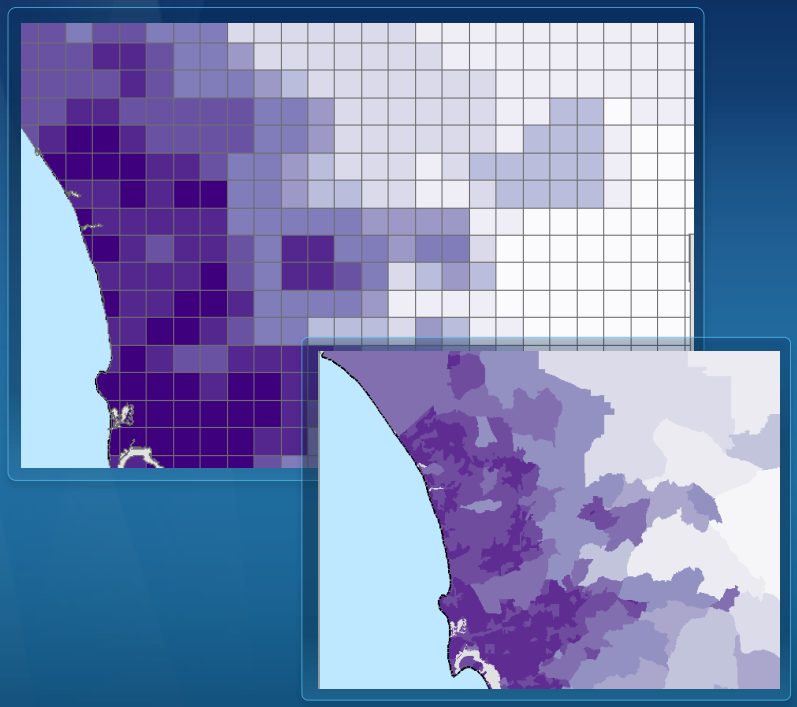
$$V_t = \sum_{s=1}^{s} \frac{V_s \cdot A_{st}}{A_s}$$

where: V_t is the value in target zone t;
V_s is the population in source zone s;
As is the area of source zone s; and
A_{ts} is the area of target zone t overlapping source zones.

Area weight:
12/21 = 0.5714

Simple area weighting





Simple area weighting

Assumptions

- Only suitable for population counts
  - not suitable for variables that are related to the areal unit in which they are located
- Assumes that the data is homogenous i.e. evenly distributed across an area
- Other approaches should be used where other informative data is available

Dasymetric mapping

- **Dasymetric mapping disaggregates area data to a finer unit of analysis**
  - **uses ancillary data to help refine population locations**

Filtered area weighting

- The use of land use/land cover as ancillary data for filtering

- Mask area weighting:
  - uses ancillary data as a filter or mask to eliminate the areas deemed to be uninhabitable e.g. water areas
- Filtered area weighting:
  - apportion data to higher resolution ancillary dataset
    - Area data e.g. buildings or land cover (percentages or regression), light emissions
    - Point data e.g. postcode data, housing centroids
    - Linear data e.g. roads

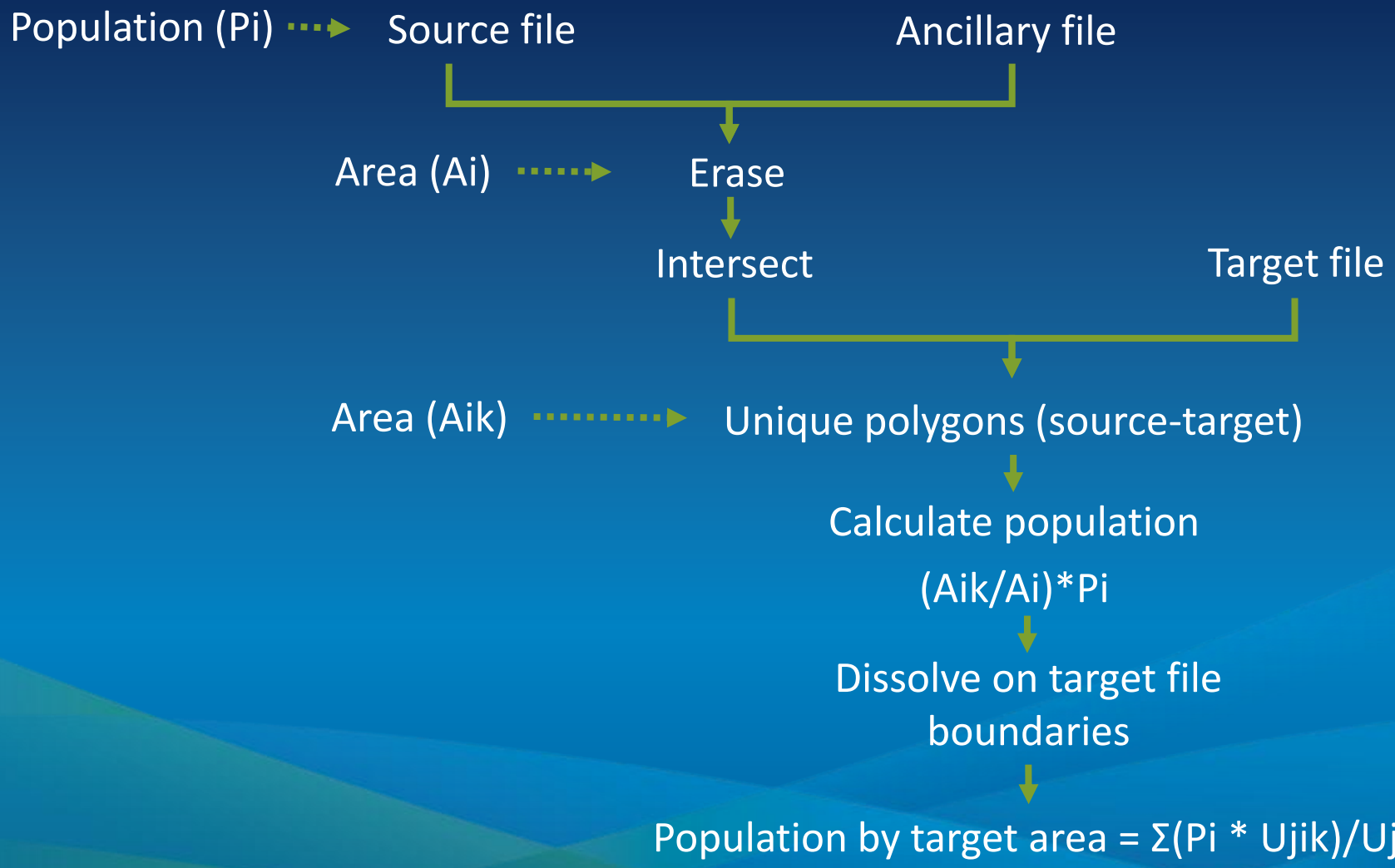
Mask mapping (Binary filtered area weighting)

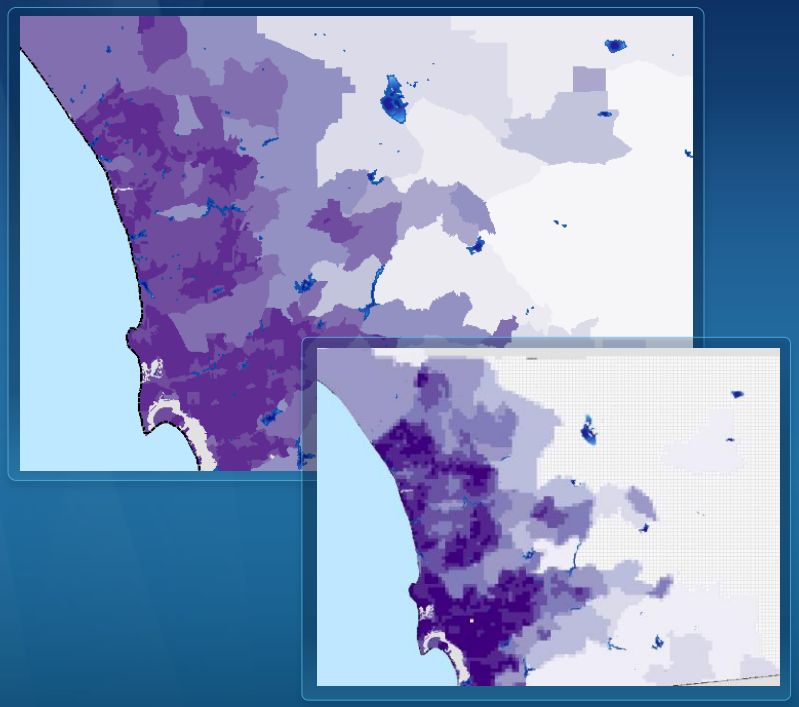
- Improvement on simple area weighting
  - Uses ancillary dataset to define areas where population will not be located e.g. water areas

$$V_t = \sum_{s=1}^{s} \frac{V_s \cdot A_{stf}}{A_{sf}}$$

where: A_{stf} is the area of overlap between target (t) and source (s) minus filtered area (f)
A_{sf} is target area minus filtered area

Mask area weighting



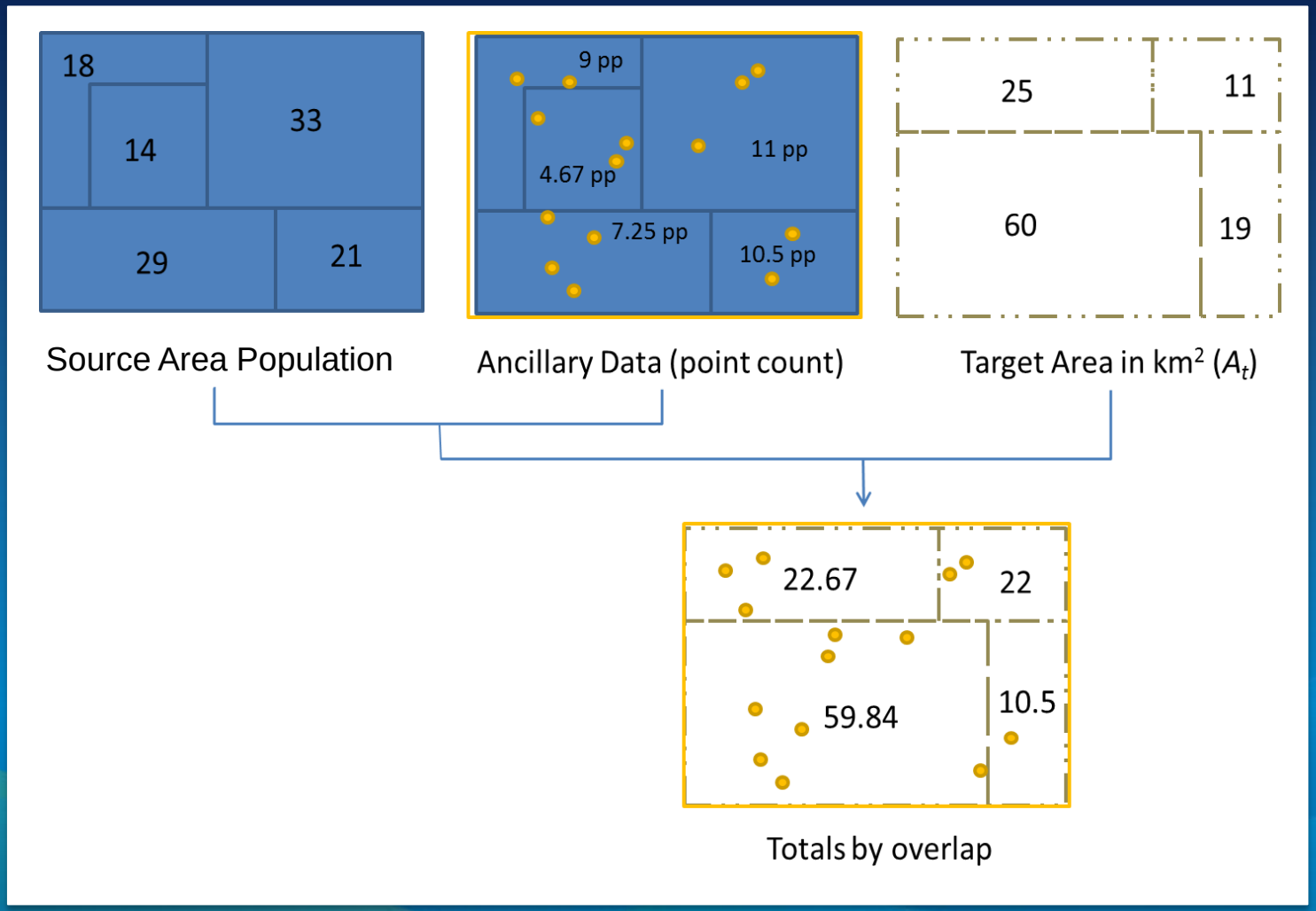


Mask area weighting

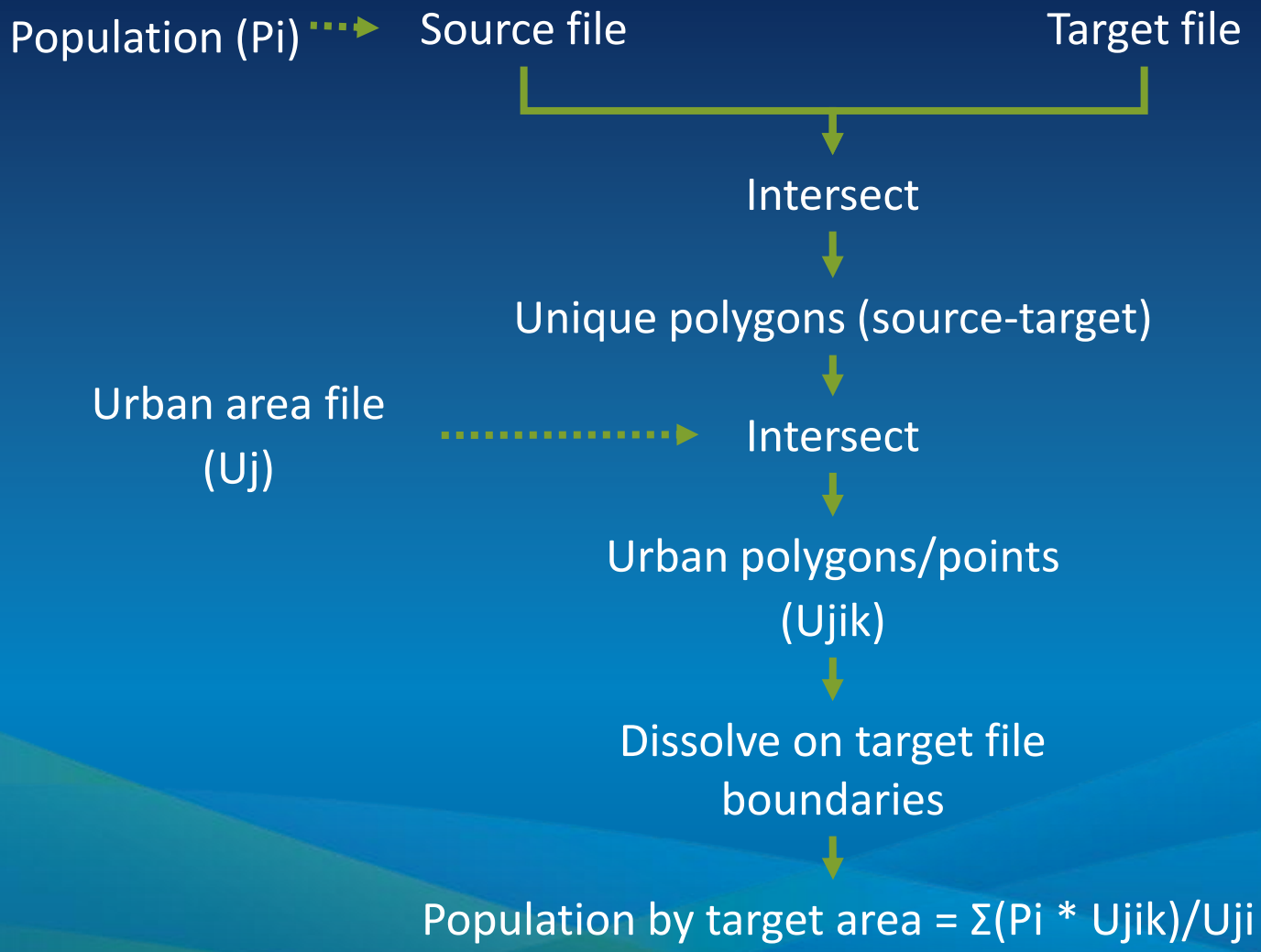
Assumptions:

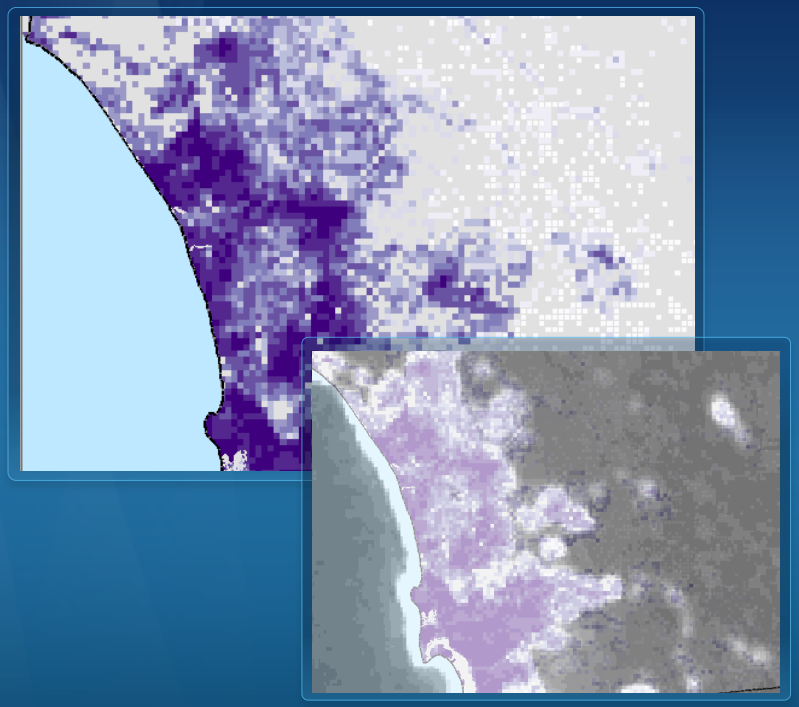
- As with simple are weighting, only suitable for population counts
- Attempts to improve on the re-distribution of data by eliminating data from selected areas
  - But still assumes homogeneity in your data

Filtered Area Weighting



Filtered area weighting





Filtered area weighting

Assumptions:

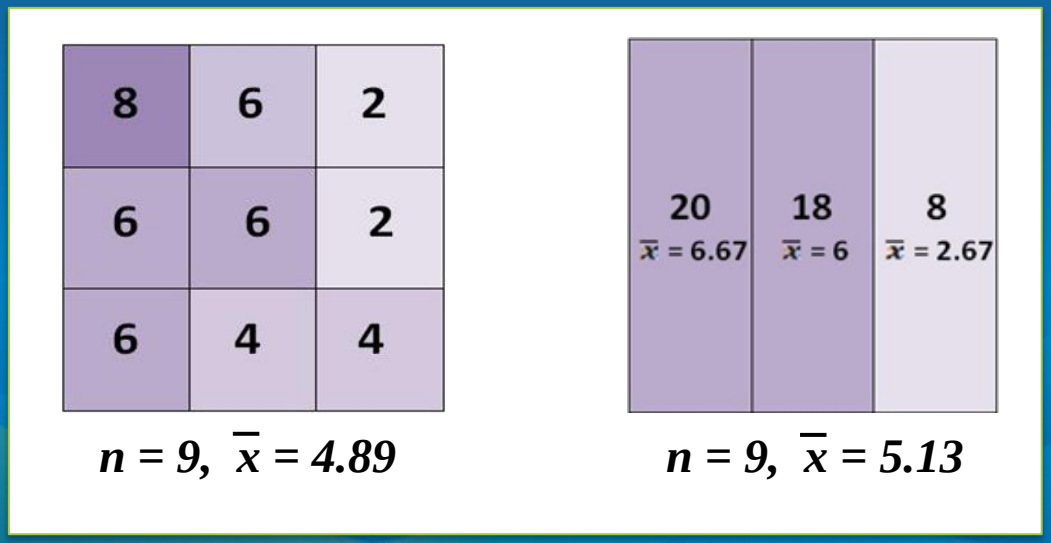
- Again, only suitable for population counts
- Improve on the re-distribution of data by using finer scale data
  - Better deals with homogeneity in your data
- Requires ancillary data
  - Adaptable method

Geostatistical Analyst: Kriging areal interpolation

- Does not assume that data values are evenly distributed across areas
- Predicted values and standard errors can be estimated
- The continuous surfaces of values can then be aggregated to any target geography
- Works with aggregated area data, count or rate data and event data
- Assumes that the polygon values reflect average values from an underlying smooth surface
- Session here in demo theater: Thurs 9 – 9.30am

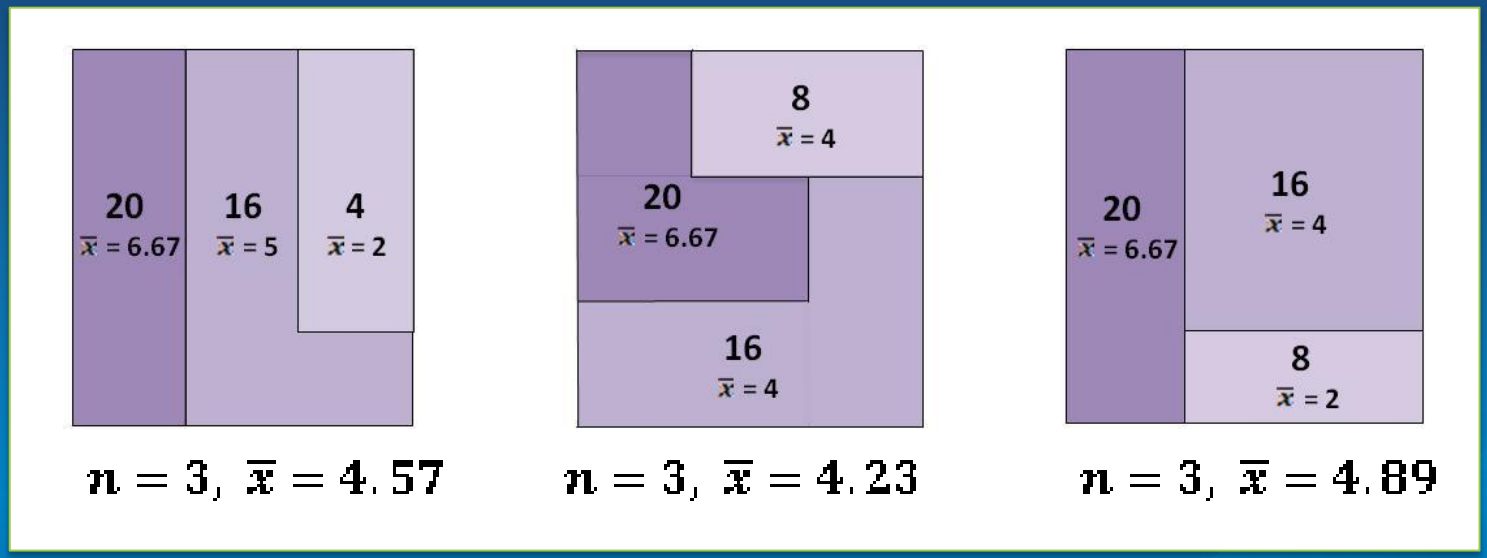
Modifiable Areal Unit Problem

- Two observed effects of MAUP, the scale problem and the aggregation or zonation issue
- The scale problem:
  - the tendency for different statistical results to be obtained when aggregating information from a set of areal units into fewer and larger zones (i.e., a smaller-scale representation)

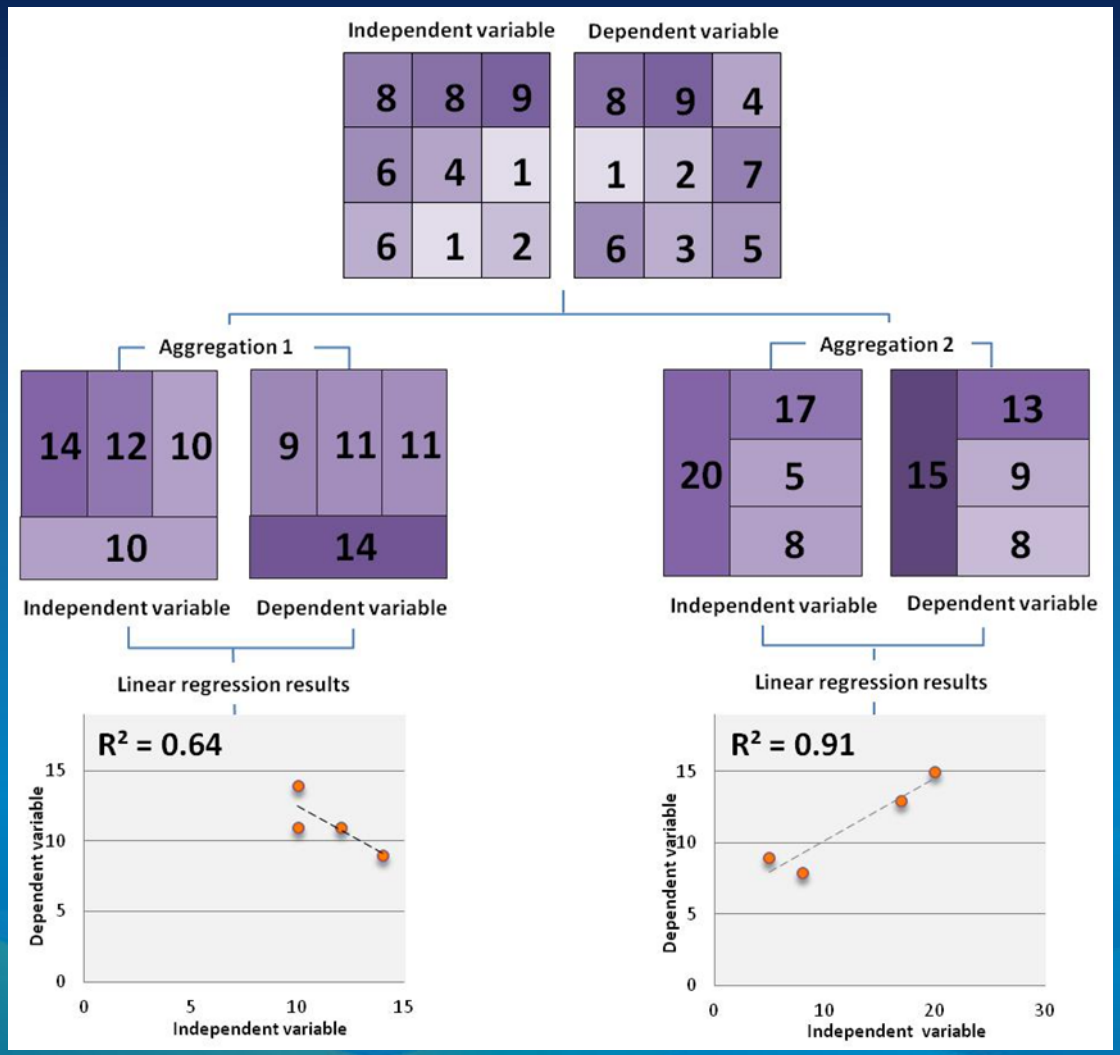


Modifiable Areal Unit Problem

- **Aggregation problem:**
  - **variability in results obtained through variations in the shape of areas**



Effects of MAUP on statistical analysis



Overall effects of MAUP

- The MAUP effect does not invalidate results
- But both the scale and aggregation (or zonation) of your data should be appropriate to your question
  - The areal framework should be meaningful in terms of the underlying processes being analyzed
  - Results only relate to the chosen scale of analysis and aggregation
- The assumptions of many aspatial statistical methods are violated by the dependencies seen within spatial data
- Perform sensitivity analysis to see how small changes in the aggregation scheme (i.e., changes in shape) impact results
  - Use ModelBuilder or a Python script as part of your workflow

Conclusion

- Given an overview of different approaches to areal interpolation:
  - Simple area weighting
  - Mask area weighting
  - Filtered area weighting
  - Outlined assumptions for each method
- Discussed the effects of MAUP

Thank you

lbeale@esri.com

esri

lbeale@esri.com