Tobler’s First Law of Geography

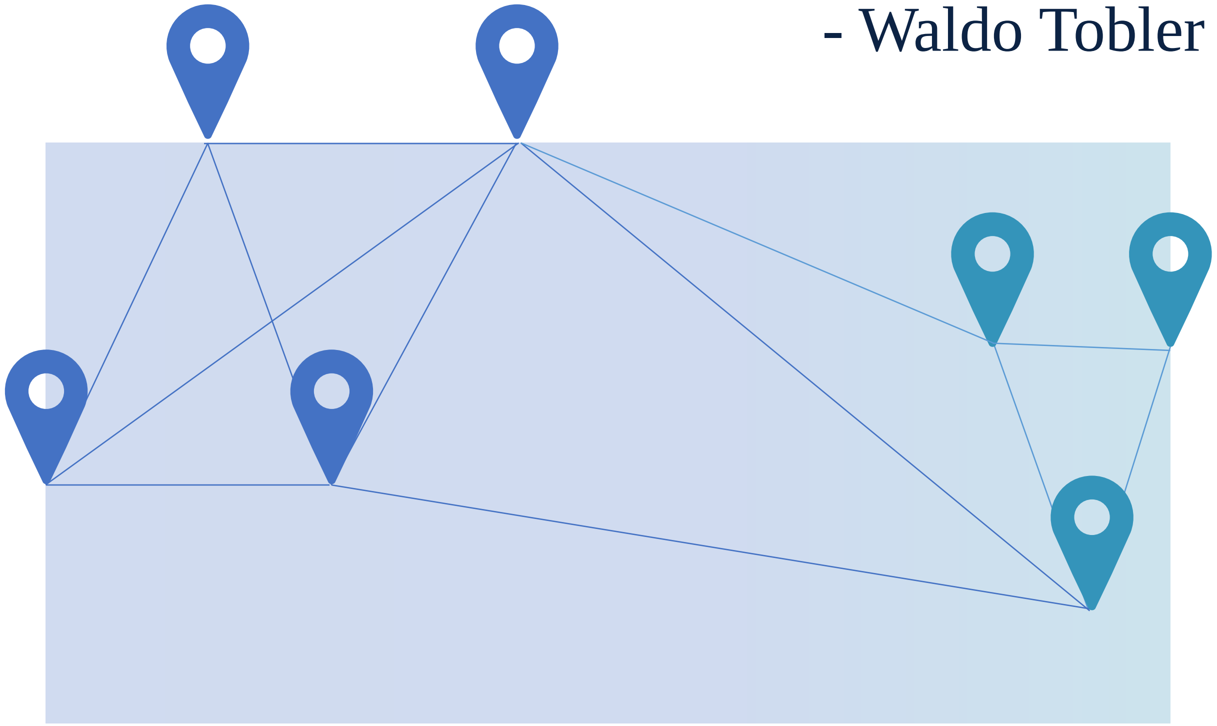
Objects near each other are more likely to be similar to one another

Objects that are distant from each other are more likely to be different from each other

This aspect of nature keeps coming up again and again in GIS!

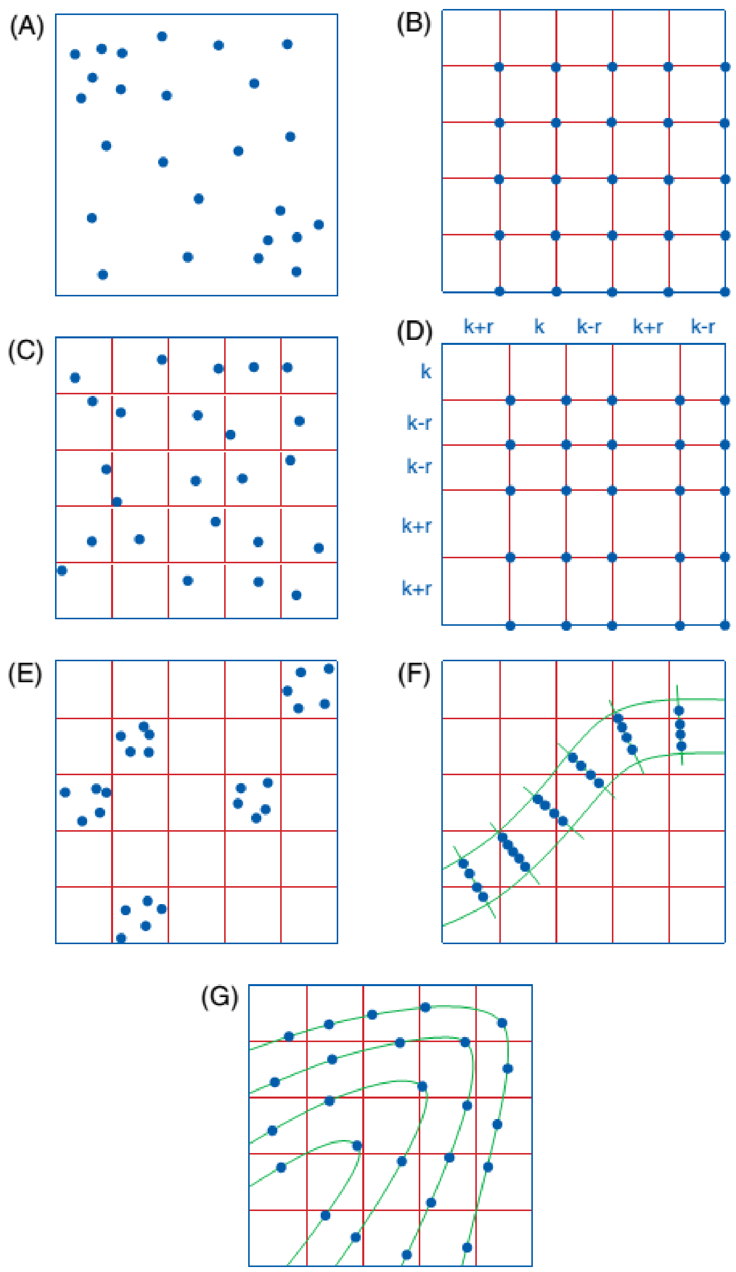
Everything is related to everything else, but near things are more related than distant things.

\- Waldo Tobler



Spatial Sampling

- You can think of sampling as the process of selecting points from within an area, called a **sample frame**
  - We select some areas within the frame, but discard most others
  - How we choose which points to keep can determine the quality of our data



Spatial Sampling

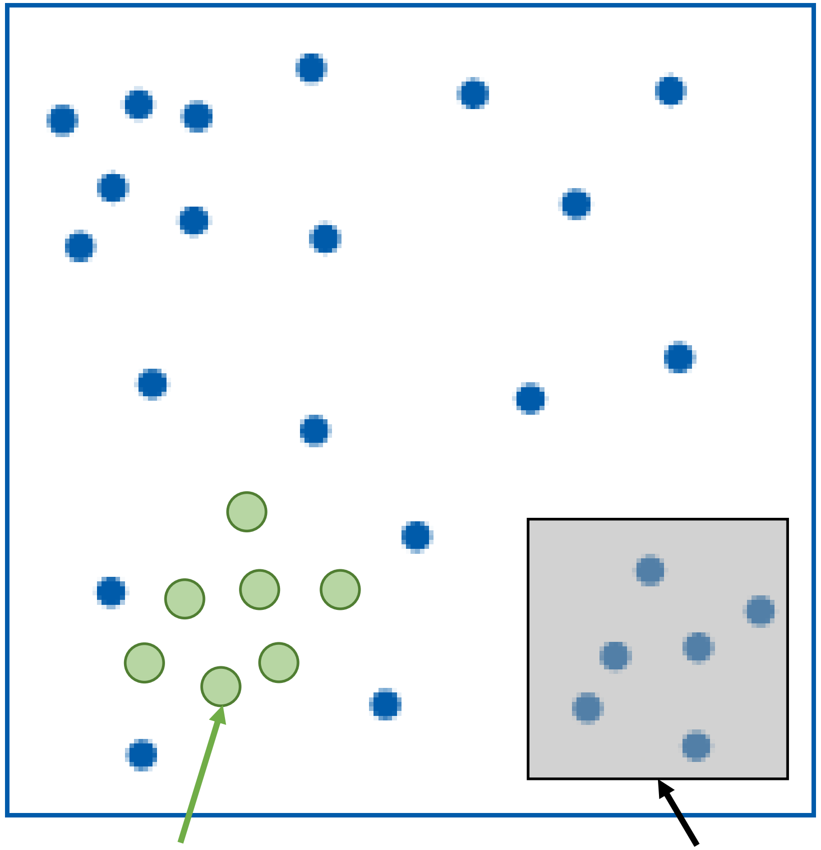
- Scientific sampling requires that each element in the sample frame have a known and pre-specified chance of selection
  - If some elements have a greater or lower chance of being selected our sample is said to be **biased**
  - If every element of interest has an equal chance of being selected our sample is said to be **unbiased**

Random Sampling

In theory, a random sample is best

- Each location has equal chance of being selected **(Unbiased)**
- Easy to do, randomly select x, y coordinates

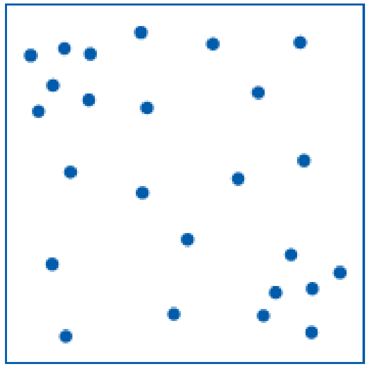
- There is a chance, that all samples will miss important features
    - Can be difficult to implement in practice



Smaller elements may be completely missed

Large elements, like buildings or crop fields, may be over sampled

(A)



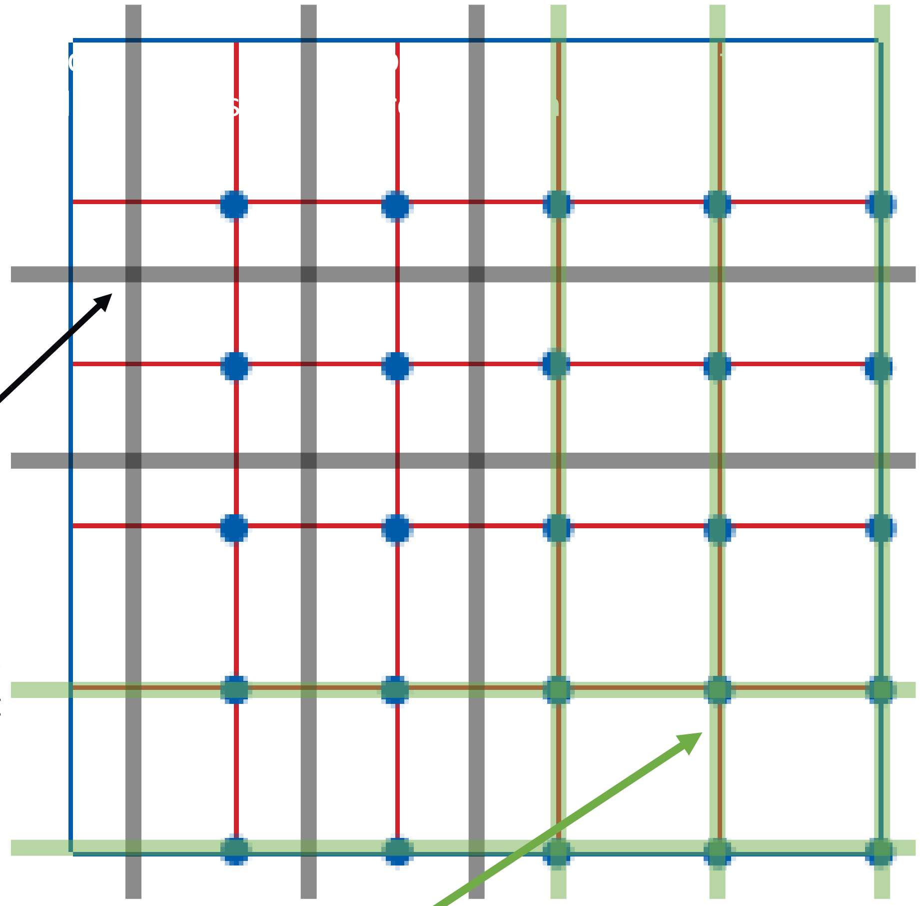
Spatial Sampling

- We have a number of approaches to account for the drawbacks of random sampling
  - Repeated sampling
    - Repeated samples should average towards the actual population values
    - More time consuming, and requires more resources
  - Systematic **(Biased)** sampling
    - Create a sample design that trades a sampling scheme for randomness
    - May miss features that have a regular pattern or cluster

Systemic Sampling

A regular grid is drawn on the sample space to ensure evenness

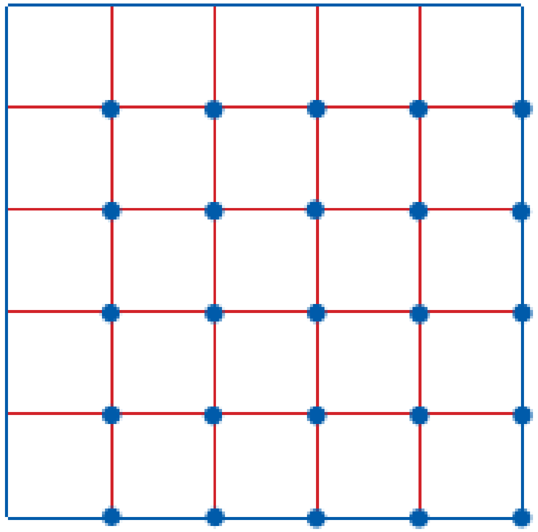
- Can be a good for areas that contain few features and abrupt boundaries
- Not ideal for attributes that exhibit periodicity, like roads



May completely miss regularly spaced features like roads that are offset from the grid

May oversample regularly spaced features like roads that line up with the grid

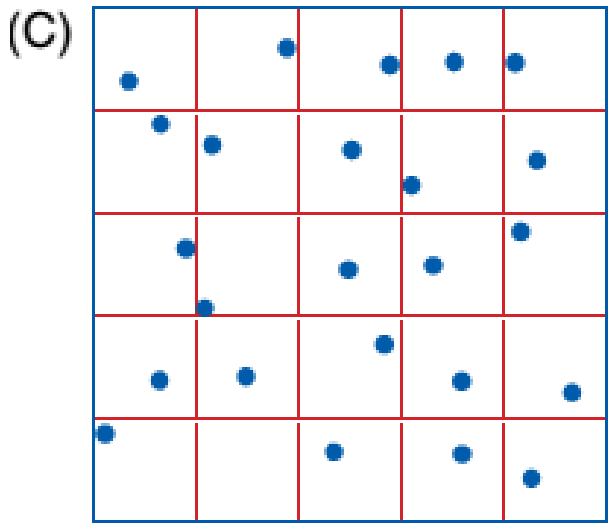
(B)

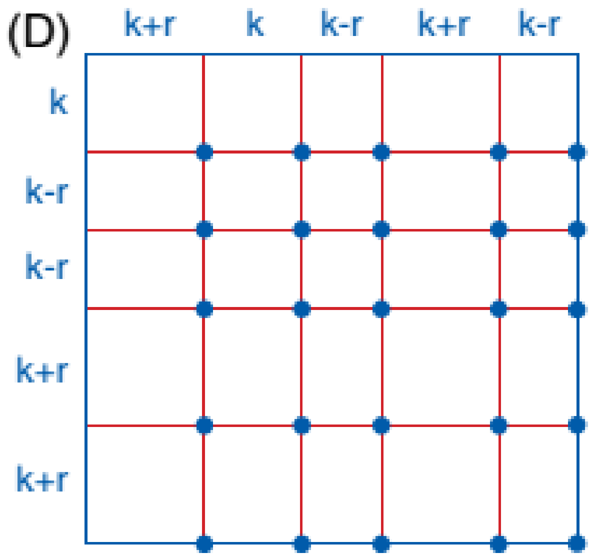


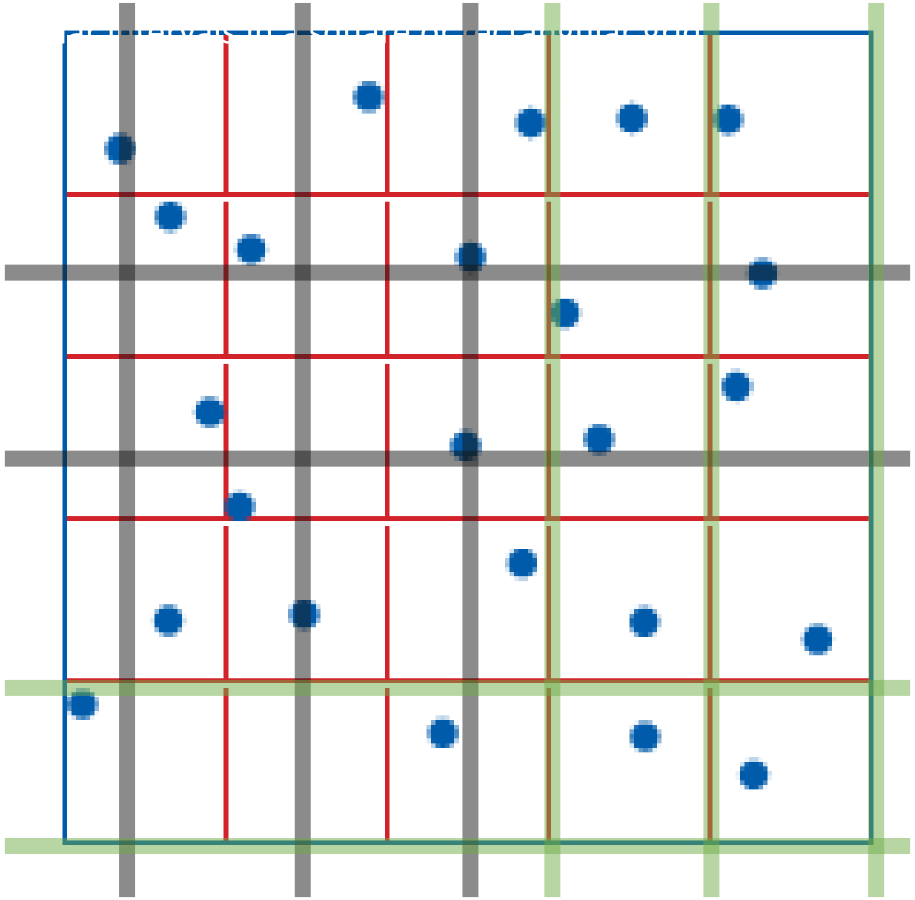
Stratified Random Sampling

Helps to address the issues with Systematic Sampling by sampling at random locations within a regularly spaced grid

- However, this can be time consuming and costly
- Alternatively use a grid, but add a random space between grid lines



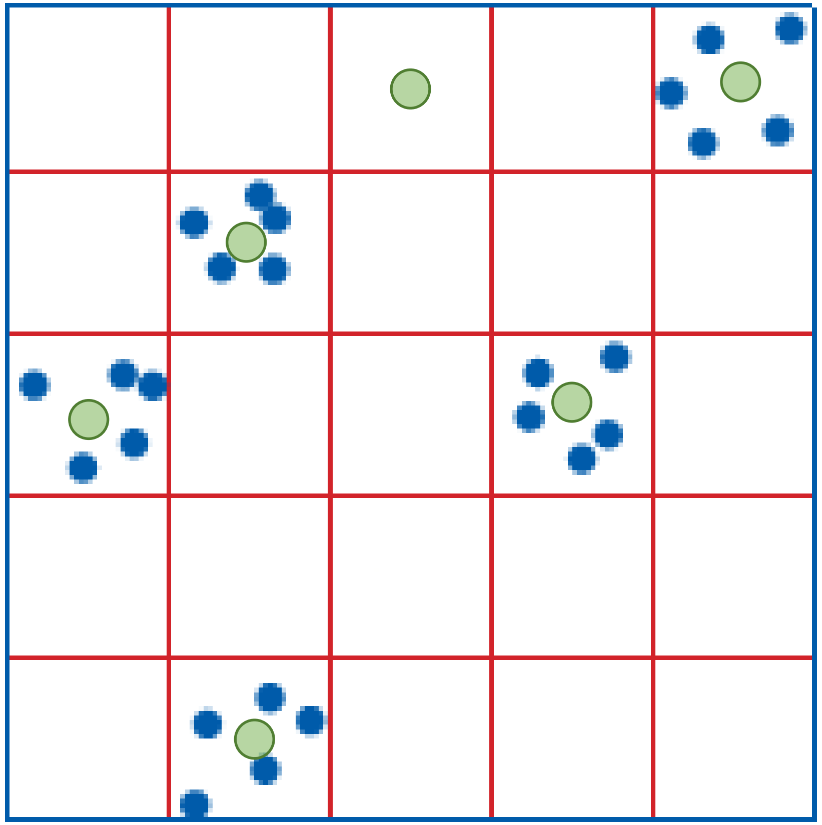




Cluster Sampling

Intense sampling of features in clusters around a number of selected locations

- Locations can be selected for specific features:
  - E.g., all shopping centers
- Or, locations can be selected at random across the grid
- Efficient, but may not be representative of the broader population you are interested in



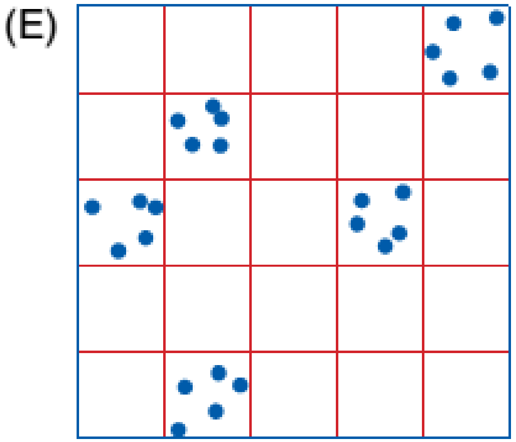
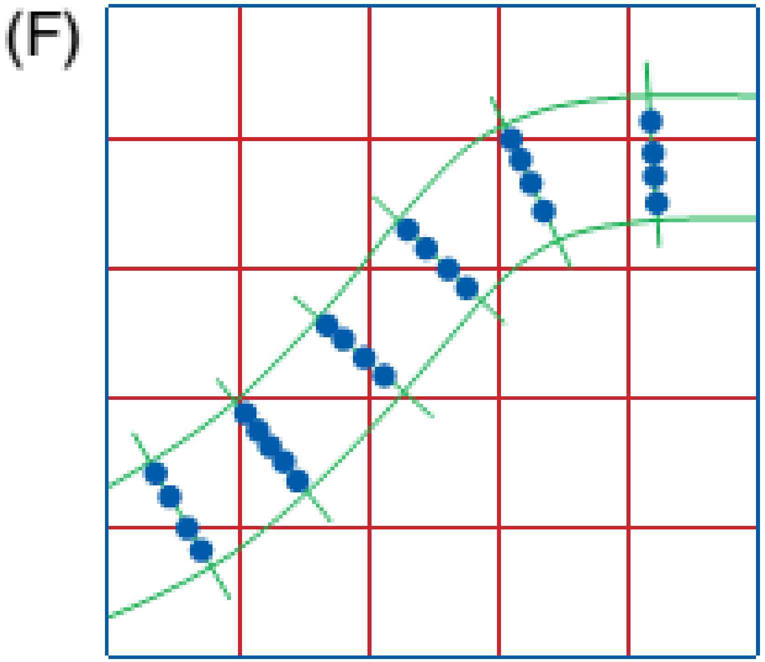


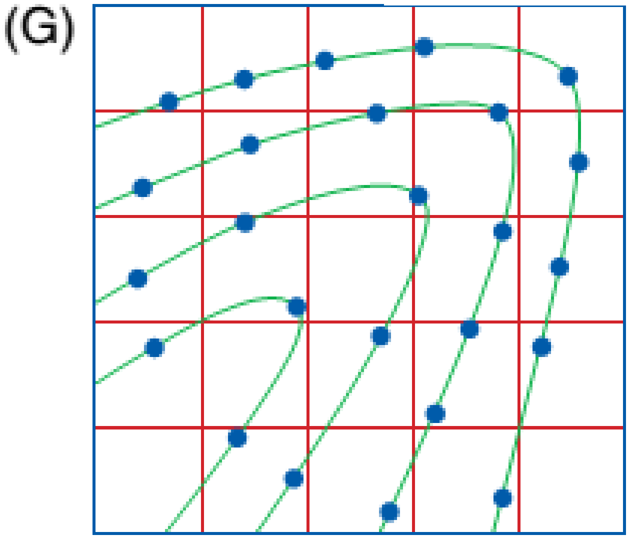
Image source: Longley et al (2015), p. 40

Transect & Contour Sampling

Allows you to focus your efforts along observed areas of variability

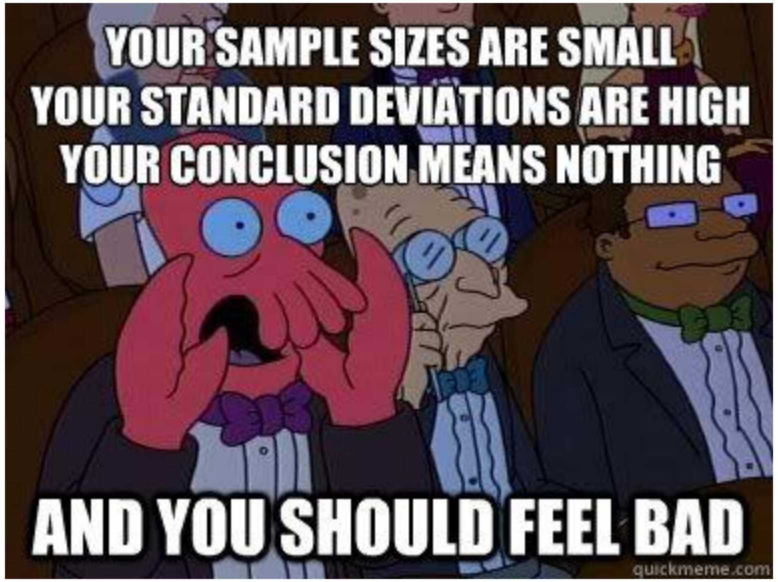
- Efficient: you only sample in the area features of interest are likely to be found
- No sampling outside of area of focus
- Requires a good understanding of spatial structure for maximum effectiveness
- Transect: Most commonly used along line features like roads, rivers
- Contour: can be used on hillslopes, drainage basins, etc.





How Many Samples?

- The **number of samples** required to adequately represent a population is a function of **how similar** units of that population are.
- Spatial structure can **vary** wildly across a landscape
    - A little knowledge of your study area will help you to establish ways to best stratify to maximize returns with minimal effort
- Need to **balance** effective coverage with the cost (in $ or time)
    - Eg. short vs. long census form



Spatial Autocorrelation

- When the value of one object is related to the value of other nearby objects
  - If you know the value of one object, you can make a reasonable guess at the value of the other object
  - That doesn’t mean that the values are the same, however!
  - It also doesn’t mean that the values of one object directly affect the values of the other (although that is possible)
    - Correlation does not imply causation!
    - Merely that there is a relationship between them, or a relationship to a third object that determines the values of the other two

Patterns of Spatial Autocorrelation

Here are some examples of different kinds of spatial autocorrelation:

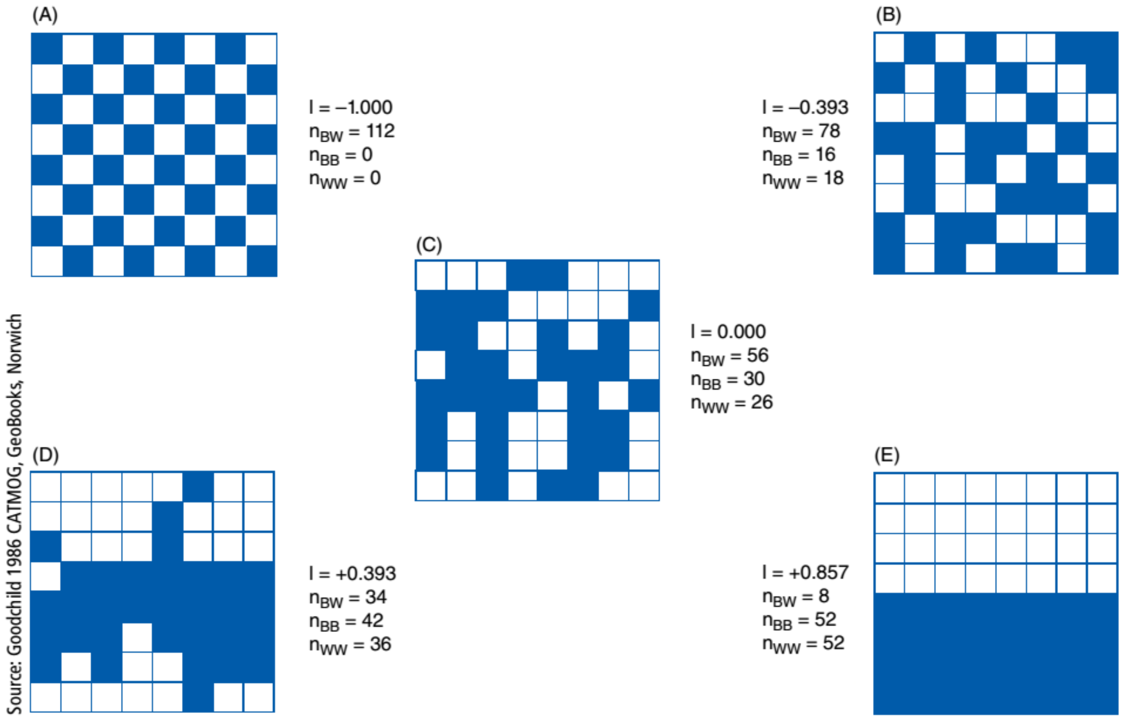
a) Extreme negative autocorrelation

b) A dispersed arrangement

c) Spatial independence

d) Spatial clustering

e) Extreme positive autocorrelation



Spatial Autocorrelation & Violating Statistical Assumptions

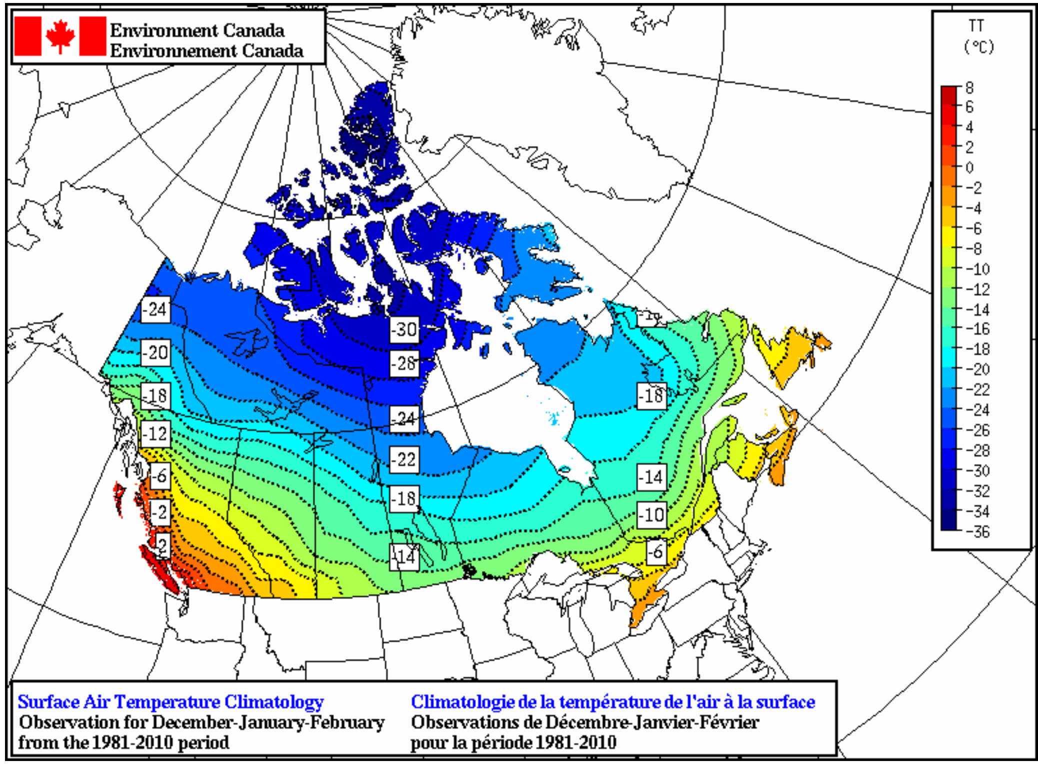
- This is a problem when it comes to spatial statistics, because most statistics assume that there is no relationship between objects
  - By violating this assumption, we “break” many common statistics!
  - For this reason, a separate field of spatial statistics explores different ways of accessing similar information, or we acknowledge this as a flaw in our assumptions
- **It is also a benefit!** We can exploit spatial autocorrelation to our advantage.

Spatial Interpolation

- The process of “filling in the blanks” that you just performed is called **interpolation**
  - Over a 2D or 3D surface we call this **spatial interpolation**
- Intelligent guesswork in which we attempt to make **reasonable estimates** of the values of a continuous field at places **where we do not have measurements**

Spatial Interpolation

- Spatial interpolation only makes sense for a **continuous field**
  - With **Nominal** (categorical) data, this can be problematic
- Works well with:
  - Rainfall, temperature, pressure
    - Estimate between weather stations
  - Elevation between measured locations
  - Changing resolutions for rasters



Spatial Interpolation & Tobler’s First Law

All spatial interpolation methods incorporate **distance to known samples**

- Sound familiar? This is Tobler’s First Law!
- Closer samples may be given more weight than distant ones
- A threshold is usually set, to determine the maximum distance to take samples from

Continuous fields **tend** to exhibit strong positive **spatial autocorrelation**, it is reasonable to assume that missing values are **likely to be similar** to those around them in the field

Inverse Distance Weighting

- One of the most common methods
    - Adjusts cells based on nearby observations
        - Weight cells by distance from observation points