Glossary

adjacent. Two areas are said to be adjacent if they share a border. Also see *contiguity matrix*.

AR(1). See *autoregressive errors*.

areal data. Areal data is a term for data on areas. SAR models are appropriate for areal and lattice data.

areas. Areas is an informal term for geographic units.

attributes. Attributes is the name given to the variables included in standard-format shapefiles.

autoregressive errors. Spatially autoregressive errors account for spatially lagged correlation of the residuals. ρ is the correlation parameter. It is not a correlation coefficient, but it shares certain properties with correlation coefficients. It is bounded by -1 and 1, and 0 has the same meaning, namely, no correlation.

autoregressive models. Spatially autoregressive models include a spatially lagged dependent variable or spatially autoregressive errors. See [SP] **Intro 1**.

balanced and strongly balanced. Panel data are balanced if each panel contains the same number of observations. They are strongly balanced if they record data for the same times (subcategory).

border and vertex. Consider the following map:

```
                    +-------+
                    |       |
          +---------+   C   |
          |    B    |       |
+---------+---------+-------+
|         A         |
+-------------------+
```

A and B share a border because there is a line segment separating them. For the same reasons, B and C share a border.

A and C share a vertex. They have only a single point in common.

How should you treat vertex-only adjacency? This issue arises when constructing a contiguity matrix. It is up to you whether a vertex in common is sufficient to label the areas as contiguous. Vertex-only adjacency occurs frequently when the shapes of the geographic units are rectangular.

choropleth map. A choropleth map is a map in which shading or coloring is used to indicate values of a variable within areas.

contiguity matrix and ex post contiguity matrix. A contiguity matrix is a symmetric matrix containing 0s and 1s before normalization, with 1s indicating that areas are adjacent.

`spmatrix create contiguity` creates contiguity matrices and other matrices that would not be considered contiguity matrices by the above definition. It can create first-order neighbor matrices containing 0s and 1s. That is a contiguity matrix. It can create first- and second-order neighbor matrices containing 0s and 1s. That is not a contiguity matrix strictly speaking. And it can create other matrices where second-order neighbors are recorded as 0.5 or any other value of your choosing.

And finally, even if the matrix started out as a contiguity matrix strictly speaking, after normalization the two values that it contains are 0 and c.

As a result, commands like `spmatrix summarize` use a different definition for contiguity matrix.

An ex post contiguity matrix is any matrix in which all values are either 0 or c, a positive constant. It is meaningful to count neighbors in such cases. Thus, the matrix W2 created by typing

```
. spmatrix create contiguity W2, second
```

is an ex post contiguity matrix, and the matrix W created by typing

```
. spmatrix create contiguity W, first second(0.5)
```

is not.

coordinate system. A coordinate system is the encoding used by numbers used to designate locations. Latitude and longitude are a coordinate system. As far as Sp is concerned, the only other coordinate system is planar. Planar coordinates are also known as rectangular or Cartesian coordinates. In theory, standard-format shapefiles provide planar coordinates. In practice, they sometimes use latitude and longitude, but standards for encoding the system used are still developing. See [SP] **spdistance** for a more complete description, and see [SP] **Intro 4** for how you can determine whether coordinates are planar or latitude and longitude.

covariate. See *explanatory variable*.

cross-sectional data. Cross-sectional data contain one observation per spatial unit. Also see *panel data*.

.dbf files. See *shapefiles*.

dependent variable. See *outcome variable*.

distance matrix. A distance matrix is a spatial weighting matrix based on some function of distance. Usually that function is 1/distance, and the matrix is then called an inverse-distance spatial weighting matrix.

explanatory variable. An explanatory variable is a variable that appears on the right-hand side of the equation used to "explain" the values of the outcome variable.

FIPS codes. FIPS stands for federal information processing standard. FIPS codes are used for designating areas of the United States. At the most detailed level is the five-digit FIPS county codes, which range from 01001 for Autauga County in Alabama to 78030 for St. Thomas Island in the Virgin Islands. The FIPS county code includes counties, U.S. possessions, and freely associated areas.

The first two digits of the five-digit code are FIPS state codes. The two-digit code covers states, U.S. possessions, and freely associated areas.

The five-digit code appears in some datasets as the two-digit state code plus a three-digit county code. The full five-digit code is formed by joining the two-digit and three-digit codes.

geographic units. Geographic units is the generic term for places or areas such as zip-code areas, census blocks, cities, counties, countries, and the like. The units do not need to be based on geography. They could be network nodes, for instance. In this manual, we also use the words places and areas for the geographic units. Also see *spatial units*.

GIS data. GIS is an acronym for geographic information system. Some of the information in shapefiles is from such systems.

ID, _ID variable. An ID variable is a variable that uniquely identifies the observations. Sp's _ID variable is an example of an ID variable that uniquely identifies the geographic units. Sp's _ID variable is a numeric variable that uniquely identifies the observations in cross-sectional data and uniquely identifies the panels in panel data.

idistance spatial weighting matrix. An idistance spatial weighting matrix is Sp jargon for an inverse-distance spatial weighting matrix.

i.i.d. I.i.d. stands for independent and identically distributed. A variable is i.i.d. when each observation of the variable has the same probability distribution as all the other observations and all are independent of one another.

imported spatial weighting matrix. An imported spatial weighting matrix is a spatial weighting matrix created with the `spmatrix import` command.

instrumental variables. Instrumental variables are variables related to the covariates (explanatory variables) and unrelated to the errors (residuals).

inverse-distance spatial weighting matrix. An inverse-distance spatial weighting matrix is a matrix in which the elements $W_{i,j}$ before normalization contain the reciprocal of the distance between places j and i. The term is also used for inverse-distance matrices in which places farther apart than a specified distance are set to 0.

lags. See *spatial lags*.

latitude and longitude. See *coordinate system*.

lattice data. Lattice data are a kind of area data. In lattice data, all places are vertices appearing on a grid. SAR models are appropriate for lattice data and areal data.

neighbors, first- and second-order. First-order neighbors share borders. Second-order neighbors are neighbors of neighbors.

normalized spatial weighting matrix. A normalized spatial weighting matrix is a spatial weighting matrix multiplied by a constant to improve numerical accuracy and to make nonexplosive autoregressive parameters bounded by -1 and 1. See *Choosing weighting matrices and their normalization* in [SP] **spregress** for details about normalization.

outcome variable (dependent variable). The outcome variable of a model is the variable appearing on the left-hand side of the equation. It is the variable being "explained" or predicted.

panel data. Panel data contain data on geographic units at various times. Each observation contains data on a geographic unit at a particular time, and thus the data contain multiple observations per geographic unit. Also see *cross-sectional data*.

places. Places is an informal term for geographic units.

planar coordinates. See *coordinate system*.

proximity matrix. Proximity matrix is another word for distance matrix.

SAR. SAR stands for spatial autoregressive or simultaneous autoregressive, which themselves mean the same thing but are used by researchers in different fields. See *autoregressive models* and *autoregressive errors*.

shapefiles. Shapefiles are files defining maps and more that you find on the web. A shapefile might be *name*`.zip`. *name*`.zip` contains *name*`.shp`, *name*`.dbf`, and files with other suffixes.

In this manual, shapefiles are also the shapefiles as described above translated into Stata format. They are Stata datasets named *name*`_shp.dta`.

To distinguish the two meanings, we refer to standard-format and Stata-format shapefiles.

Sp. Sp stands for spatial and refers to the SAR system described in this manual.

Sp data. Sp data are data that have been `spset`, whether directly or indirectly. You can type `spset` without arguments to determine whether your data are `spset`.

spatial lags. Spatial lags are the spatial analogy of time-series lags. In time series, the lag of x_t is x_{t-1}. In spatial analysis, the lag of x_i—x in place i—is a weighted sum of x in nearby places given by $\mathbf{Wx}$. See [SP] **Intro 1**.

spatial units. Spatial units is the term we use for the units measuring distance when the coordinates are planar. For instance, New York and Boston might be recorded in planar units as being at $(\_CX, \_CY) = (1.3, 7.836)$ and $(1.447, 7.118)$. In that case, the distance between them is 0.0284 spatial units. Because they are about 190 miles apart, evidently a spatial unit is 6,690 miles. Also see [SP] **spdistance**.

spatial weighting matrix. A spatial weighting matrix is square matrix $\mathbf{W}$. $\mathbf{Wx}$ plays the same role in spatial analysis that L.$\mathbf{x}$ plays in time-series analysis. One can think of $\mathbf{W}$'s elements as recording the potential spillover for place j to i.

Spatial weighting matrices have zero on the diagonal and nonzero or zero values elsewhere. A contiguity spatial weighting matrix would have 0s and 1s. $W_{i,j} = W_{j,i}$ would equal 1 when i and j were neighbors.

The scale in which the elements of spatial weighting matrices are recorded is irrelevant. See [SP] **Intro 2**.

spatially autoregressive errors. See *autoregressive errors*.

spillover effects. Spillover effects and potential spillover effects are the informal words we use to describe the elements of a spatial weighting matrix. $W_{i,j}$ records the (potential) spillover from place j to i. See [SP] **Intro 2**.

standard-format shapefile. See *shapefiles*.

Stata-format shapefile. See *shapefiles*.

strongly balanced. See *balanced and strongly balanced*.

time variable. The time variable is the variable in panel data that identifies the second level of the panel. The variable is not required to measure time, but it usually does.

user-defined matrix. A user-defined matrix is a spatial weighting matrix created by typing

```
spmatrix userdefined
spmatrix fromdata
spmatrix spfrommata
```

vertex. See *border and vertex*.

For suggested citations, see the FAQ on citing Stata documentation.