

# DIVA-GIS --- Exercise 1

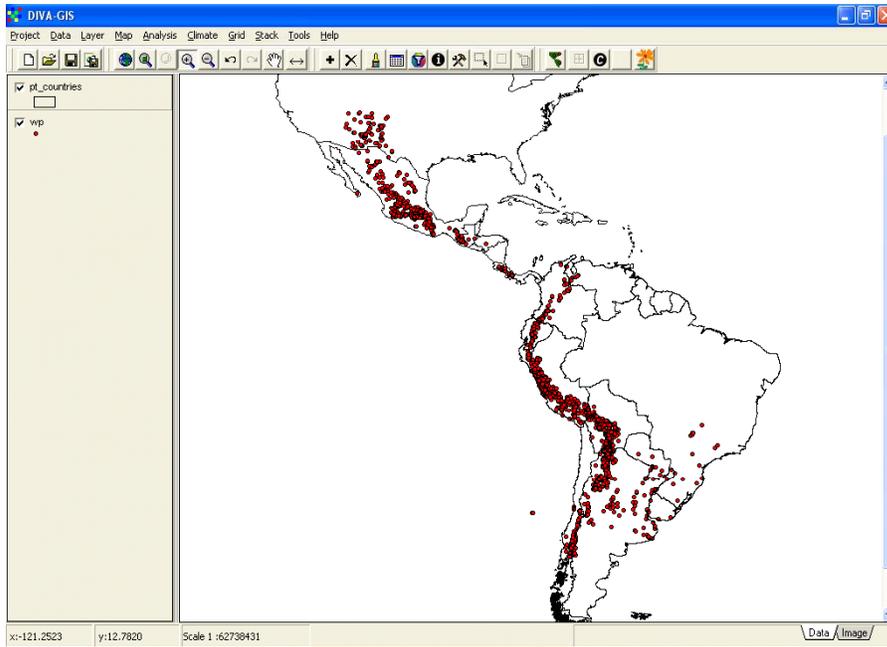
## Distribution and diversity of wild potatoes

You should have gone through the tutorial and at least glanced through the manual before starting this exercise.

Wild potatoes (Solanaceae; *Solanum* sect. *Petota*) are relatives of the cultivated potato. There are nearly 200 different species that occur in the Americas. Their geographic distribution was recently described by Hijmans and Spooner (Hijmans, R.J., and D.M. Spooner, 2001. Geographic distribution of wild potato species. *American Journal of Botany* 88:2101-2112). In this exercise, we analyze some of the same data that were used in this paper.

### A. Import data

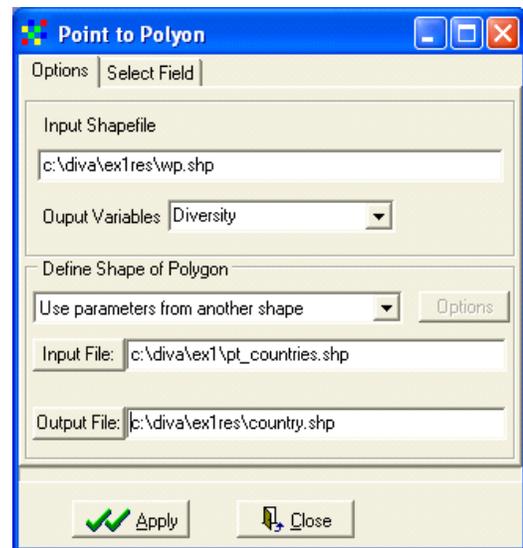
- 1) Create a folder called "diva-ex1". Download the data for this exercise, put them in the new folder and unzip the files. Open the file 'wildpot.txt' with Excel to convert the data in degrees and decimal minutes to decimal degrees (see Chapter 2 of the manual), and save the file in tab delimited text format. Make sure that you do not lose the decimal numbers of the coordinates as you save the file.
- 2) Use this text file to make a shapefile called "wild potatoes", save it in the data folder and add it to the map. Also add the 'pt\_countries' shapefile to the map. Your map should now look like this:



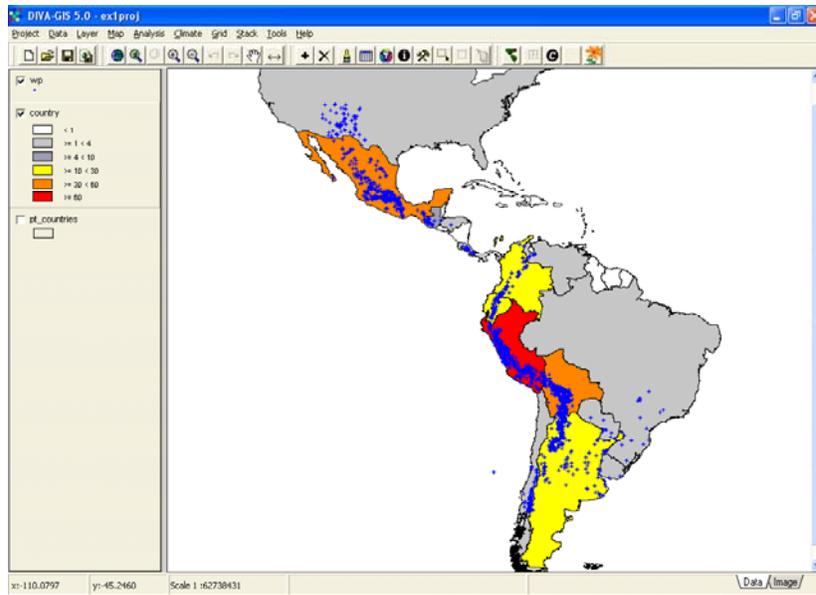
## B. Summarize by country

- 3) We are first going to summarize the data by country. Make the potato shapefile the active file and click on *Analysis/Point to Polygon*. Select 'species' as the field of interest. Add the shapefile of countries in the 'define shape of polygon' box. Select an output filename and press *Apply*.

The result is a new countries layer. Make this layer visible and the other two layers invisible. Double click on the new layer and change its legend attributes.



First, on the *single* tab, double click on the symbol and change its style to 'solid fill'. Then go to the *classes* tab, select the SPPfield, 6 classes, and reset the legend. SPP stands for the number of distinct species. Look in the manual to find out what the other fields are. After that, manually change the values of the different classes to get something like this:



This map shows that Peru is the country with most potato species, followed by Bolivia and Mexico (which country has more?), and then by Argentina, Colombia and Ecuador.

- 4) If you open the table associated with the shapefile, you will see that there are records for many countries that have no potatoes at all. To simplify our table, we first select the records with more than 0 observations. Using *Layer/Select Records* and the *select by query* tab.

After entering the appropriate query all countries with potatoes will be colored yellow on the map (they are 'selected'). Now save the selection to a new shapefile, add it to the map, and open the table. The table should look something like this:

COUNTRY	DIVAD	OBS	SPP	MARGALEF	MENHINICK	SHANNON	SIMPSON
ARGENTINA	1	1473	28	3.7	0.7	0.001	11.94
BOLIVIA	5	985	38	5.37	1.18	0.03	19.7
BRASIL	6	17	3	0.71	0.48	0.24	3.09
CHILE	7	94	3	0.44	0.21	0.58	2.06
COLOMBIA	8	104	13	2.58	1.18	0.04	6.66
COSTA RICA	9	24	1	0	0	1	1
ECUADOR	13	139	14	2.64	1.1	0.15	7.28
GUATEMALA	17	58	6	1.23	0.66	0.02	4.05
HONDURAS	21	1	1	0	0	1	-1
MEXICO	24	846	36	5.19	1.2	0.02	15.42
PANAMA	26	13	2	0.39	0.28	0.46	2.17
PARAGUAY	27	19	2	0.34	0.23	0.05	1.12
PERU	28	1042	89	12.66	2.73	0.004	20.74
UNITED STATES	37	157	3	0.4	0.16	0.01	2.01
URUGUAY	38	4	2	0.72	0.5	0.5	3
VENEZUELA	39	14	3	0.76	0.54	0.14	2.84

Record: 1 of 16

Statistic Highlight Pan To Zoom To

The information by country is of interest and can be useful. However, it is not very helpful for understanding the spatial distribution of wild potatoes

because the countries have such different sizes and shapes, and because most are very large. It is in most cases better to use grids with cells of equal area, and that is what we will do next.

Save the project to, e.g., 'exercise1A'.

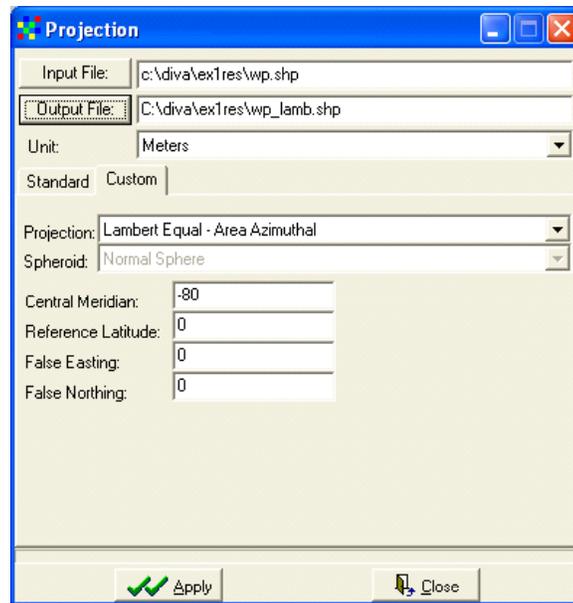
### C. Project the data

To be able to use a grid with cells of equal area, the data needs to be projected. If the lat/long data were used, cells of say 1 square degree would get smaller as you move away from the equator: think of the meridians (vertical lines) on the globe getting closer to each other as you go towards the poles.

For small areas, UTM would be a good projection, but in this case we will use a projection that can be used for a complete hemisphere: the Lambert Equal Area Azimuthal projection. Before you project your data, you must choose a map origin for your data. This should be somewhere in the center of your points, to minimize the distance (and hence distortion) from any point to the origin. In this case, the center could be (-80, 0).

- 5) Remove all layers from the map, except the wild potato and the original country shapefile.

Project these files using *Tools/Projection*. Choose the Lambert Equal Area Azimuthal (equatorial) projection. On the *custom* tab, change the central meridian to -80. Save the files with filenames such as 'wild potatoes lambert' and 'countries lambert'. Press *Apply*. Be patient...



Make all layers visible. If you zoom to the maximum extent, you see the projected data. Note that the shape of the countries is much more similar to their shape on a globe than before projecting. In the bottom left corner you can see that the coordinate system has changed. You likely see very large numbers now: these express the distance from the origin (-80,0 degrees) in meters. If you make the projected potato file invisible, you will see that the unprojected data is still present on the map, near the origin of the projected data. Zoom in to that area.

Clearly one cannot combine projected and unprojected data (or data in two different projections) on a single map.

- 6) Now go to *Map/Properties* and change the projection to "Other" and the units to 'meters'. This will allow for displaying a correct scale on the map.
- 7) Save the project to 'Exercise 1B'.

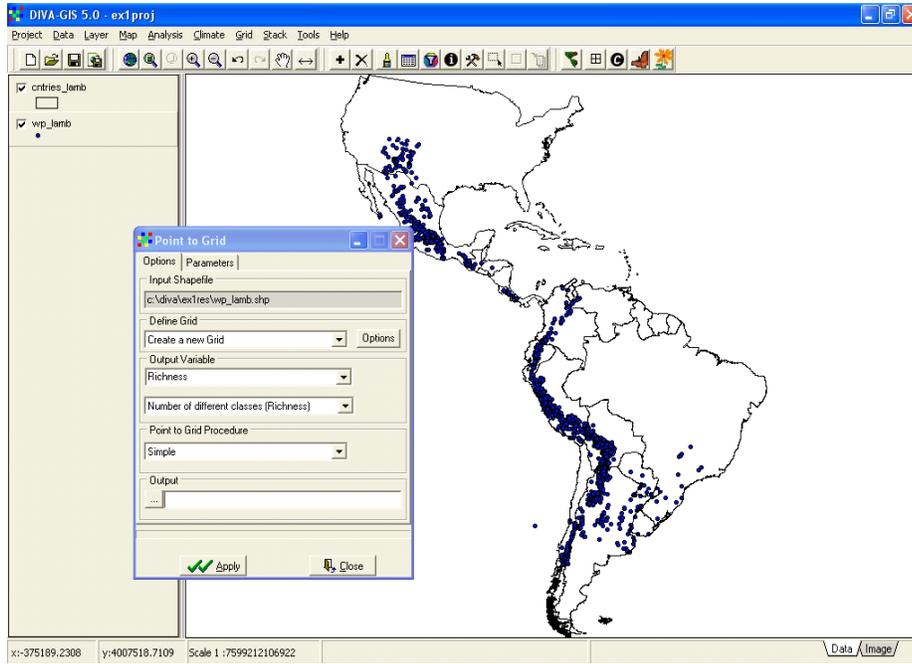
#### D - Species Richness on a grid

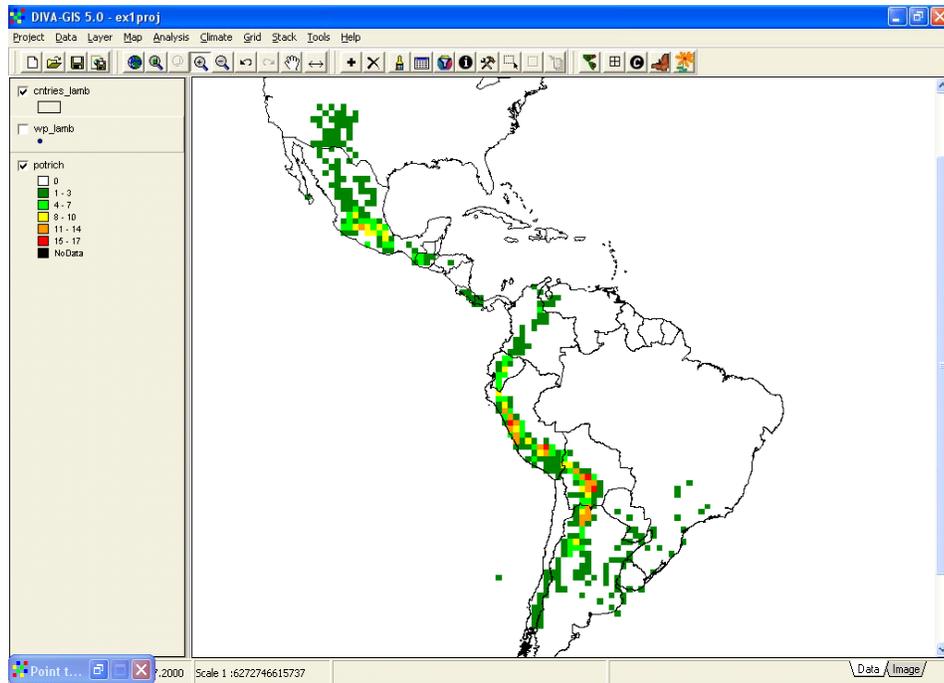
- 8) Let's determine the distribution of species richness using a grid. This can be done using the point-to-polygon option that we used before, but in most cases it is more appropriate to use *Analysis/Point to grid*.

Select 'Richness' and 'Number of different classes', and select the 'Species' Field.

Create a new grid. In the *Options* window, set the X and Y resolution to 100,000 (as the projection is in meters this means that the cells will be 100 by 100 km). Use the default option ('simple') for the Point-to-grid procedure.

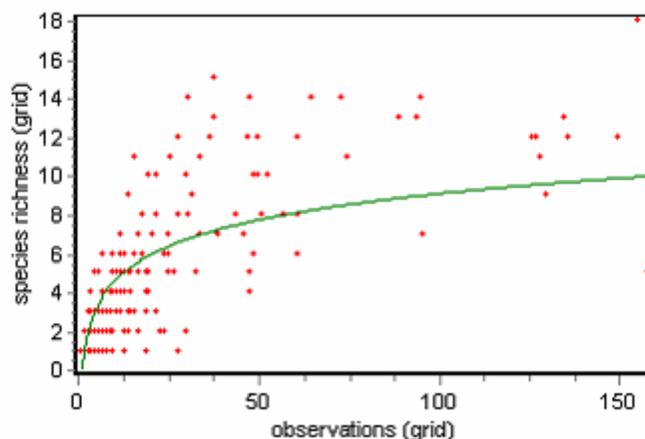
Choose an output filename (e.g., 'species richness', and press *Apply*).





9) Now make a grid of the 'Number of observations' instead of species richness. Define the new grid with the option 'Use parameters from another grid' and select the species richness grid you just made. In this way you assure that you use exactly the same grid (cell size, number of rows and columns, and geographic origin).

10) Compare the two grids with *Analysis/Regression*. Is the number of species in a cell a function of the number of observations?



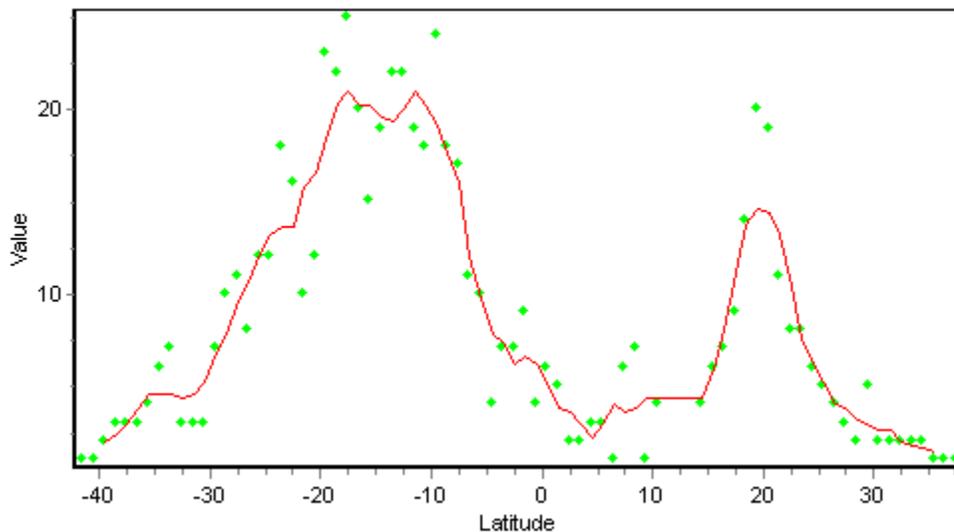
As you can see, there is a relation between the number of observations and species. Whether the number of species is, therefore, an artifact of collecting bias remains to be seen.

The problem is that this association will always exist. When there are only few species in an area, collectors will not continue to go there to increase the number of (redundant) observations.

In this case, the coefficient of determination ( $r^2$ ) is not very high (0.67) for this type of relationship. Also, there is a clear pattern in species richness maps, it is not characterized by sudden random like changes in richness. The data have spatial autocorrelation according to the Geary index, but not the Moran index. See for yourself under *Analysis/Autocorrelation*.

Additional ways to look at this 'collector-bias' problem include the use of 'rarefaction', 'richness estimators' (see manual) and predictive modeling (see Exercise 2).

- 11) There are often gradients of species richness over latitude and altitude. An easy way to investigate a latitudinal gradient in species richness is by making grid of a single column. In the grid options window (in the Point to Grid dialog) use 'adjust with resolution' and set the number of columns to 1. Then use *Grid/Transect* to create a figure like the one below. The distribution of species richness has two peaks. What would explain the low species richness between -5 and 15 degrees?



Now try an alternative way to make a graph. Export the data in the grid to a text file (*Layer/Export gridfile*) and import that file into Excel.