

Exercise One: Creating A Presence-Absence Raster Data Layer For A Species From Survey Data

In the species distribution modelling (SDM) example which you will work through in this supplementary workbook, you will be measuring the species distribution by comparing the locations where a particular species, the harbour porpoise, was recorded as present during the surveys and the locations which were surveyed but where the species was not recorded. This will be done by creating what is known as a presence-absence raster data layer in a GIS project. During this process, you will make a number of key decisions that will have major implications as to whether your SDM project will succeed or not. This includes making decisions about the resolution at which you will sample your species distribution and defining the extent of your study area. You may also make decisions about whether to include or exclude certain subsets of data in your analysis to limit the potential for false absences within your data or to account for seasonal and/or spatial variations in the niche occupied by your target species. These decisions should not be taken lightly and may require that you not only know a lot about the ecology of the species you are modelling and about your specific study area, but also that you do a lot of background reading to ensure you understand the implications of any decisions you make at this stage on the statistical analysis you use to create your SDM. For the purposes of this exercise, these decisions have been made for you, but you should think carefully before applying them to your own data or research project.

Before you start this exercise, you will first need to create a new folder on your C: drive called GIS_EXERCISES_3. To do this on a computer with the Windows 7 operating system, click on the START menu and select MY COMPUTER. In the window which opens, double-click on the icon for your C: drive (this may be called WINDOWS (C:)). This will open a window which displays the contents of your C: drive. To create a new

folder, right click on this window and select NEW> FOLDER. This will create a new folder. Now call this folder GIS_EXERCISES_3 by typing this into the folder name to replace what it is currently called (which will most likely be NEW FOLDER). This folder, which has the address C:\GIS_EXERCISES_3 will be used to store all files and data for the exercises in this book.

Next, you need to download the source files for three data sets from www.gisinecology.com/books/data/marinebiologysupplementaryworkbook. When you download these files, save them into the folder C:\GIS_EXERCISES_3 which you have just created. These data sets are:

1. Harbour_Porpoise.shp: This is a point data layer and is the shapefile of harbour porpoise locational records. You will need to download all the files called Harbour_Porpoise (each with a different extension) in order to be able to use it. Each point represents a record for a single group of porpoises recorded during a set of surveys. It is in the geographic projection and is based on the WGS 1984 datum.

2. Survey_Tracks_North_Sea.shp: This is a line data layer which contains information on survey effort during which the harbour porpoise locational data in Harbour_Porpoise.shp were recorded. You will need to download all the files called Survey_Tracks_North_Sea (each with a different extension) in order to be able to use it. Thus, this data layer tells you which locations were surveyed in this study. It is in the geographic projection and is based on the WGS 1984 datum.

3. Depth_North_Sea: This is a line data layer which contains information on water depth which you can use to create your map. You will need to download all the files called Depth_North_Sea (each with a different extension) in order to be able to use it. This data layer is in the geographic projection and uses the WGS 1984 datum. **NOTE:** The depth values are not accurate and are only provided here as an approximation for illustration purposes. As a result, they should not be used for any purpose other than these exercises.

4. Land_North_Sea: This is a polygon data layer which contains information on land in this region which you can use to create your map. You will need to download all the files called Land_North_Sea (each with a different extension) in order to be able to use it. This data layer is in the geographic projection and uses the WGS 1984 datum. **NOTE:** The land information is not accurate and is only provided here as an approximation for illustration purposes. As a result, it should not be used for any purpose other than these exercises.

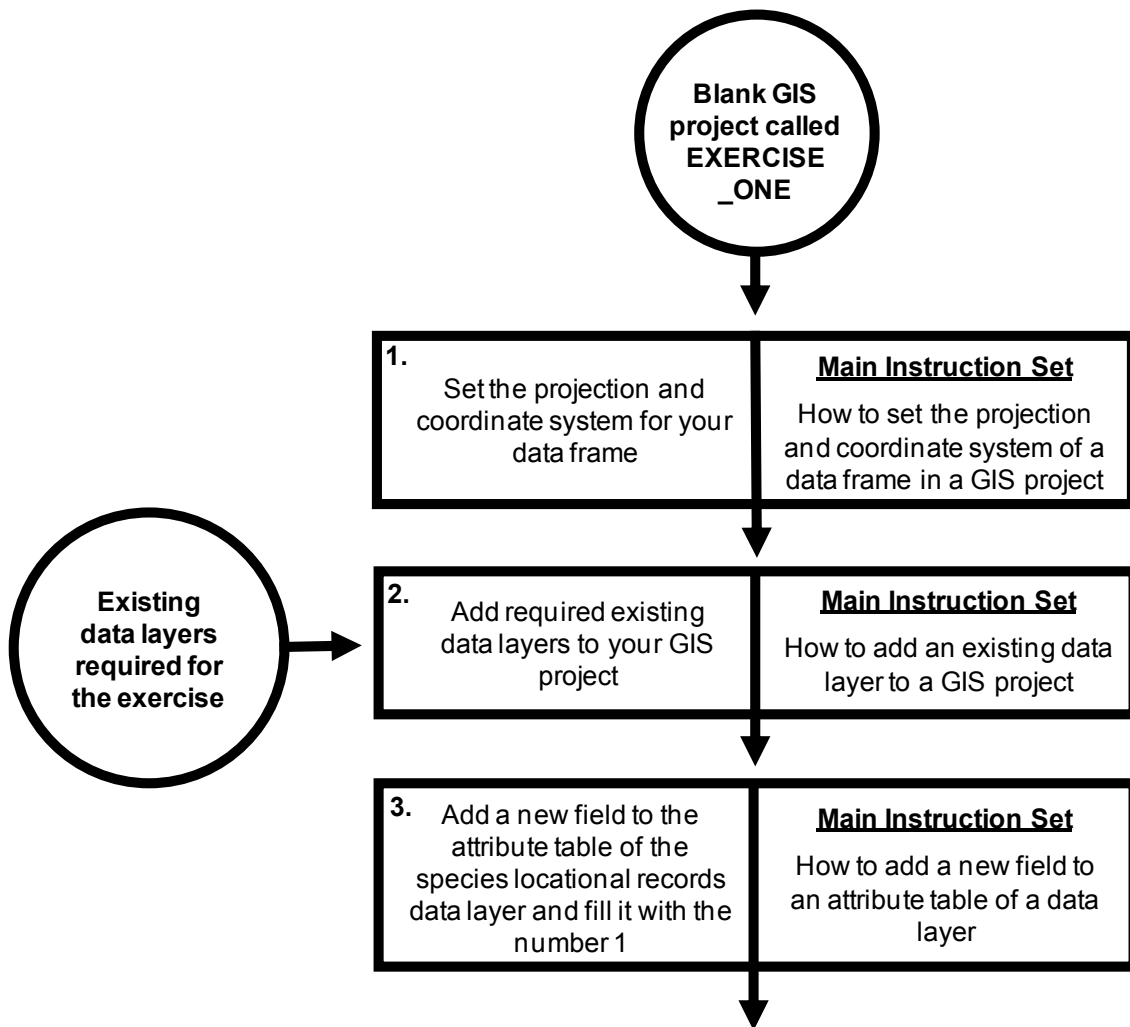
NOTE: The starting point for this exercise assumes that you have already successfully created GIS data layers from your survey data. If you are trying to adapt these instructions for your own SDM project and you have not already done this, you can find instructions which may help you create such data layers from your own data in the ‘How To ...’ section of *An Introduction To Using GIS In Marine Biology*.

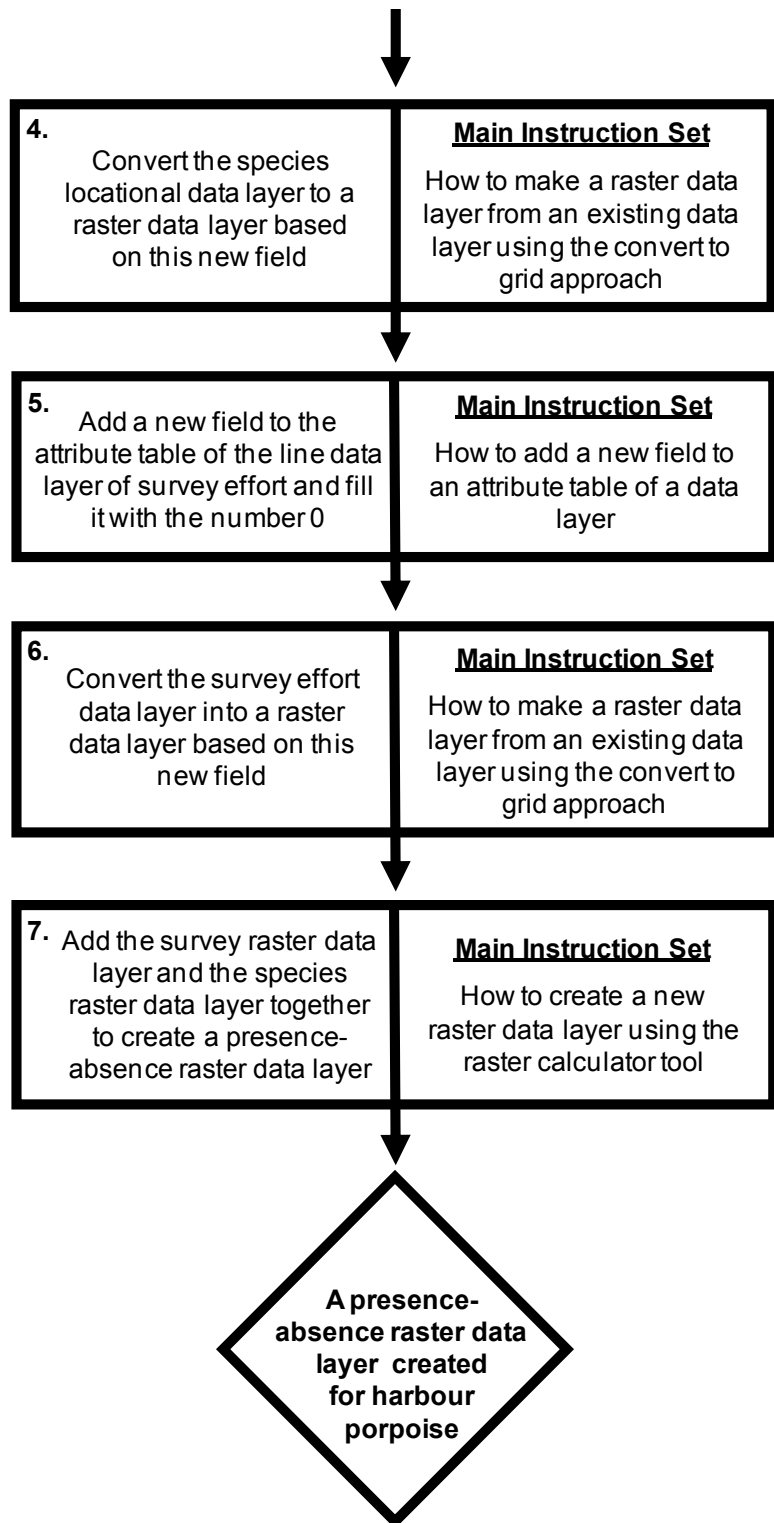
NOTE: If you wish to model observed abundance (such as the number recorded per unit of survey effort) rather than presence-absence, you can find instructions on how to create a suitable data layer of observed abundance in chapter four of *An Introduction To Using GIS In Marine Biology: Supplementary Workbook One*.

Once you have all these files downloaded into the correct folder on your computer, and understand what is contained within each file, you can move onto creating your presence-absence raster data layer. The starting point for this is a blank GIS project. To create a blank GIS project, first, start the ArcGIS software by opening the ArcMap module. When it opens, you will be presented with a window which has the heading ARCMAP – GETTING STARTED. In this window, you can either select an existing GIS project to work on, or create a new one. To create a new blank GIS project, click on NEW MAPS in the directory tree on the left hand side and then select BLANK MAP in the right hand section of the window. Now, click OK at the bottom of this window. This will open a new blank GIS project. (**NOTE:** If this window does not appear, you can start a new project by clicking on FILE from the main menu bar area, and selecting NEW. When the NEW DOCUMENT window opens, select NEW MAPS and then BLANK MAP in the order outlined above).

Once you have opened your new GIS project, the first thing you need to do is save it under a new, and meaningful, name. To do this, click on FILE from the main menu bar area, and select SAVE AS. Save it as EXERCISE_ONE in the folder C:\GIS_EXERCISES_3. The next step is to read over the summary flow diagram for this exercise. This will help you identify all the major steps you need to carry out and the individual instruction sets you will need to follow to do them. The summary flow diagram for this exercise is provided on the next two pages.

Summary Flow Diagram For Creating A Presence-Absence Raster Data layer





Once you have familiarised yourself with this summary flow diagram, you need to read through the instruction set for the first step in the flow diagram in its entirety, before working through it.

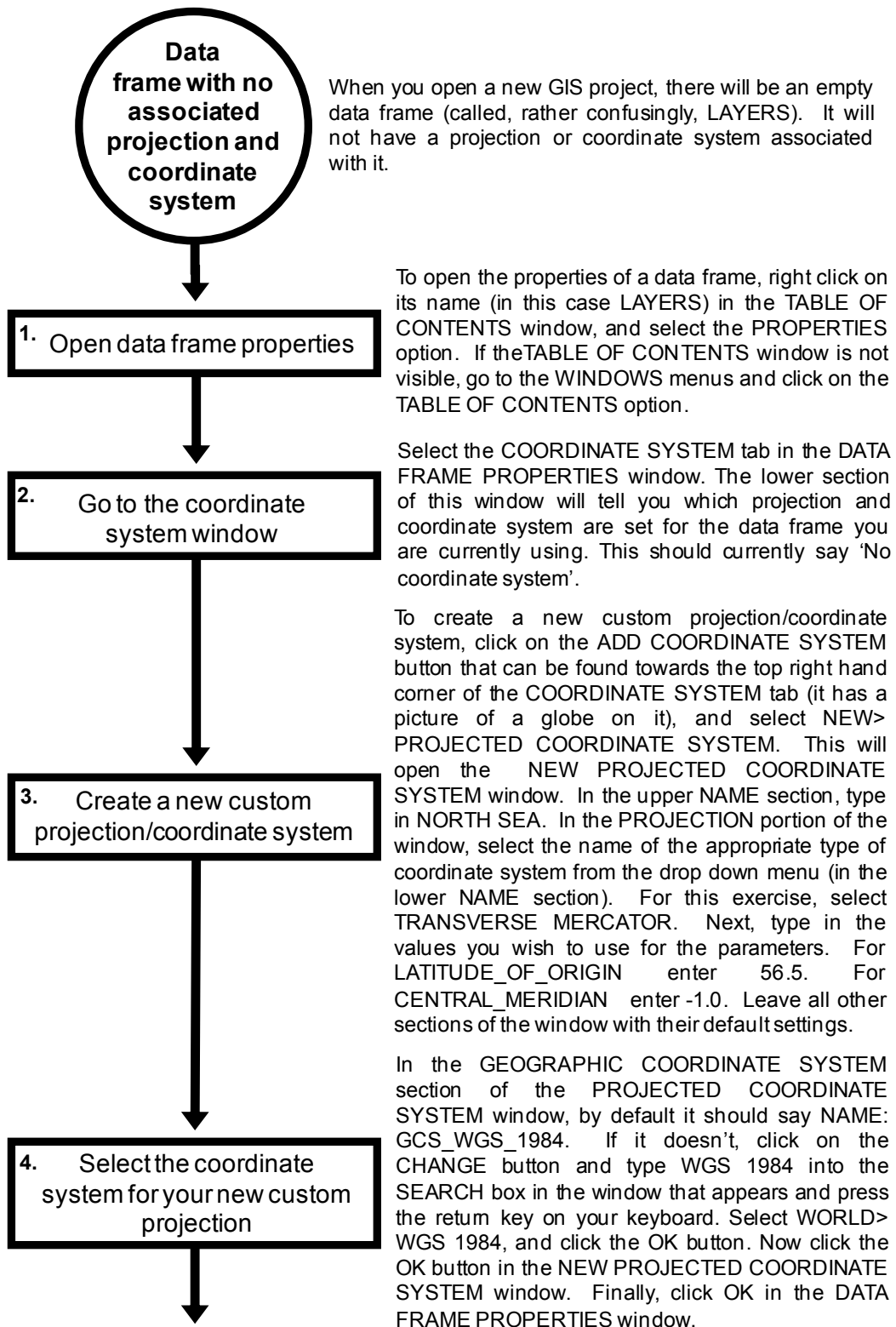
When you have completed the first major step, read through the instructions again and make sure that you have completed it properly. Next, move onto the second major step and repeat this process, and so on until you have completed all the steps in the summary flow diagram.

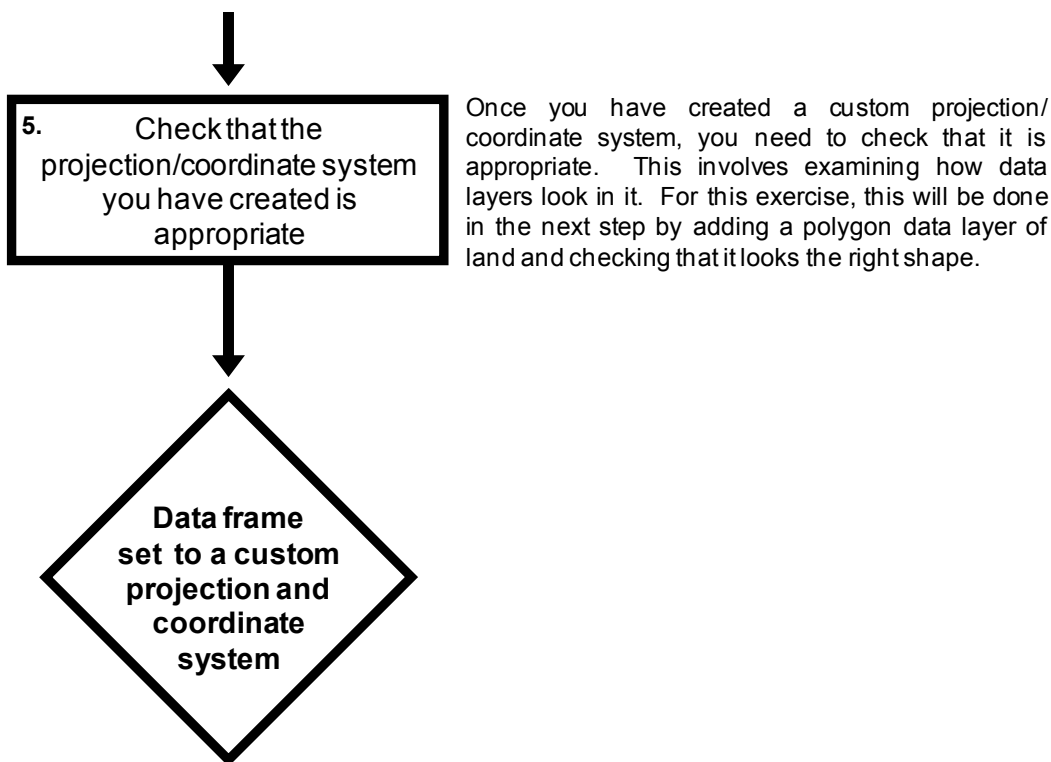
At various points, images of the contents of the MAP window, the TABLE OF CONTENTS window and/or the TABLE window will be provided so that you have an idea of what your GIS project should look like at specific points as you progress through this exercise.

Instruction Sets For The Individual Steps Identified In The Summary Flow Diagram:

STEP 1: SET THE PROJECTION AND COORDINATE SYSTEM OF YOUR DATA FRAME:

These instructions are based on the instruction set *How to set the projection and coordinate system for a data frame in a GIS project* (from *An Introduction To Using GIS In Marine Biology*). For this exercise, the projection and coordinate system will be set to a custom transverse mercator centred on the middle of the area of data collection. This central position is latitude 56.5°N and longitude 1.0°W (or latitude 56.5° and longitude -1.0° in decimal degrees). This projection was chosen as it provides an accurate representation of the shape and relative positions of features in the study area. In addition, it can also be used to accurately measure distances and areas in real world units, which will be important for later exercises. This custom transverse mercator projection will use the WGS 1984 datum. This datum is used because it is the same as the one used during data collection and for the existing data layers which you will use during the exercises in this workbook.





To check that you have done this step properly, right click on the name of your data frame (LAYERS) in the TABLE OF CONTENTS window and select PROPERTIES. Click on the COORDINATE SYSTEM tab of the DATA FRAME PROPERTIES window and make sure that the contents of the CURRENT COORDINATE SYSTEM section of the window has the following text at the top of it:

North Sea
Authority: Custom
Projection: Transverse_Mercator
False_Easting: 0.0
False_Northing: 0.0
Central_Meridian: -1.0
Scale_Factor: 1.0
Latitude_Of_Origin: 56.5
Linear Unit: Meter (1.0)

If it does not, you will need to repeat this step until you have assigned the correct projection/coordinate system to your data frame.

NOTE: When conducting a species distribution modelling project in a GIS environment, selecting an appropriate projection/coordinate system for your GIS project is a critical decision. If you select an inappropriate projection/coordinate system, you may find that you cannot create appropriate raster data layers of species occurrence and/or important environmental variables, such as slope. This means you need to think carefully about what projection/coordinate system will be appropriate for your particular study before you create your initial GIS project. See chapter four in *An Introduction To Using GIS In Marine Biology* for advice on selecting an appropriate projection/coordinate system for a specific project.

STEP 2: ADD REQUIRED EXISTING DATA LAYERS TO YOUR GIS PROJECT:

For the exercises in this book, you will need to add four existing data layers to your GIS project. These are HARBOUR_PORPOISE, SURVEY_TRACKS_NORTH_SEA, LAND_NORTH_SEA and DEPTH_NORTH_SEA. However, these four data layers are not in the same projection/coordinate system as your data frame. Instead, they are in the geographic ‘projection’, although both are based on the same WGS 1984 datum. This means the data layers will need to be transformed into the projection/coordinate system which has been chosen for this SDM project. There are two ways of doing this. The first is to add your data layers to your GIS project and then transform each one into the required projection/coordinate system. This is the way used in other supplementary workbooks in this series. However, for these exercises, you will use a different way which transforms the projection/coordinate system of a data layer as part of the process of adding it to your GIS project. This is done using the PROJECT tool as outlined in the flow diagram which starts on the next page.

NOTE: This workbook was written for ArcGIS 10.2. In ArcGIS 10.2.1 (released in early 2014), the PROJECT tool used on this step has been moved from the location described in stage one of the instructions on the next page to DATA MANAGEMENT TOOLS> PROJECTIONS AND TRANSFORMATIONS> PROJECT (i.e. it is no longer in the FEATURES sub-folder). If you cannot find this tool in the location described, try looking in this alternative location.

Existing data layer which you wish to add to your GIS project

1. Add your first data layer using the PROJECT tool

In the TOOLBOX window, select DATA MANAGEMENT TOOLS> PROJECTIONS AND TRANSFORMATIONS> FEATURE> PROJECT. This will open the PROJECT window. Select the first data layer you wish to add to your GIS project (in this case DEPTH_NORTH_SEA) by clicking on the browse button at the right hand end of the INPUT DATASET OR FEATURE CLASS window and selecting the required data layer before clicking the ADD button. In the OUTPUT DATASET OR FEATURE CLASS window enter C:\GIS_EXERCISES_3\DEPTH_NORTH_SEA_PROJECT.

Next, click on the button at the end of the OUTPUT COORDINATE SYSTEM section of the window to open the SPATIAL REFERENCE PROPERTIES window. To create the custom projection/coordinate system being used for this exercise, click on the ADD COORDINATE SYSTEM button towards the top right hand corner of the SPATIAL REFERENCE PROPERTIES window and select NEW> PROJECTED COORDINATE SYSTEM. This will open the NEW PROJECTED COORDINATE SYSTEM window. In the NAME section, type in NORTH SEA. In the PROJECTION portion of the window, select TRANSVERSE MERCATOR from the drop down menu. Next type 56.5 for LATITUDE_OF_ORIGIN and -1.0 for CENTRAL_MERIDIAN.

In the GEOGRAPHIC COORDINATE SYSTEM section of the NEW PROJECTED COORDINATE SYSTEM window, by default it should say NAME: GCS_WGS_1984. If it doesn't, click on the CHANGE button and type WGS 1984 into the SEARCH box in the window that appears and press the return key on your keyboard. Select WORLD> WGS 1984, and click the OK button. Now click the OK button in the NEW PROJECTED COORDINATE SYSTEM window. Next click the OK button in the SPATIAL REFERENCE PROPERTIES window. Finally, click OK in the PROJECT window.

You will now be asked if you wish to add the resulting data layer to your GIS project. Click YES and your data layer should appear listed in your TABLE OF CONTENTS window. **NOTE:** If you are not asked this question, or if your data layer does not automatically appear in the TABLE OF CONTENTS window, right click on the name of your data frame (LAYERS) and select ADD DATA. In the ADD DATA window, browse to C:\GIS_EXERCISES_3 and select the projected version of your data layer (in this case DEPTH_NORTH_SEA_PROJECT).

2. Check the projection/coordinate system for the data layer

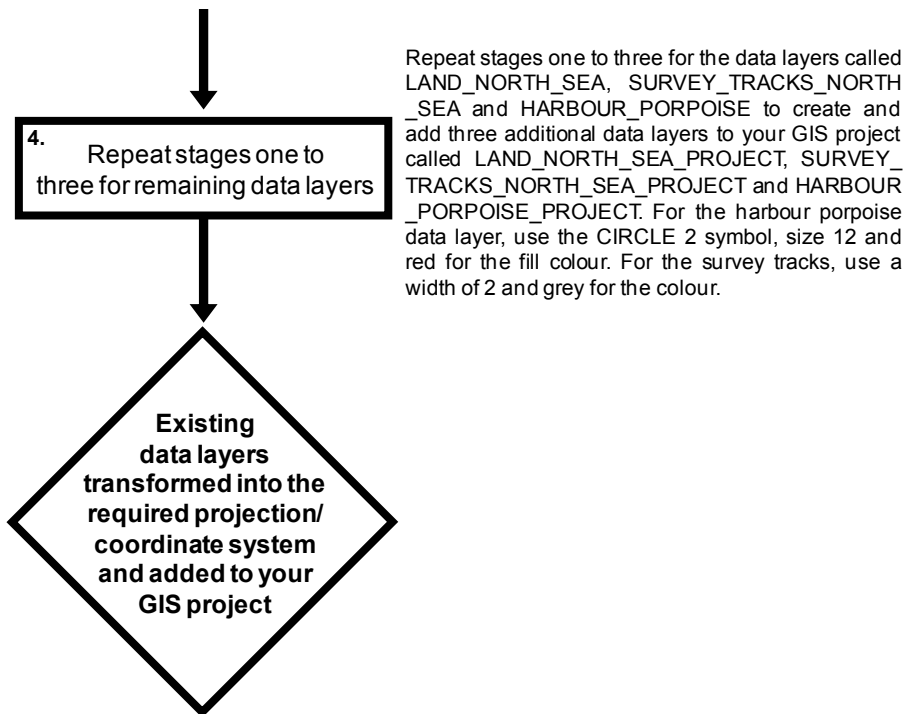
Whenever you add a data layer to a GIS project, you should always check that it has a projection/coordinate system assigned to it, and look at what this projection/coordinate system is. This is so that you know whether you will need to assign a projection/coordinate system to it, or transform it into a different projection/coordinate system before you can use it in your GIS project. To check the projection/coordinate system of your newly added data layer, in the TOOLBOX window, go to DATA MANAGEMENT TOOLS> PROJECTIONS AND TRANSFORMATIONS> DEFINE PROJECTION. This will open the DEFINE PROJECTION window. Select DEPTH_NORTH_SEA_PROJECT from the drop down menu in the top section of this window. Check that the projection/coordinate system in the lower section is called NORTH SEA. Finally, click the CANCEL button to close this window.

3. Select how the depth data layer will be displayed

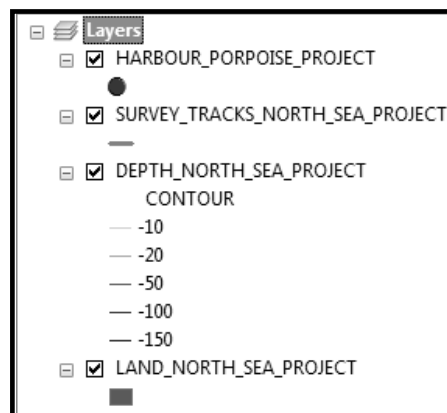
Right click on the name of the DEPTH_NORTH_SEA_PROJECT data layer in the TABLE OF CONTENTS window, and select PROPERTIES. In the LAYER PROPERTIES window, click on the SYMBOLOGY tab. On the left-hand side under show, select CATEGORIES> UNIQUE VALUES. Next, under VALUE FIELD select CONTOUR. Next, click on the ADD VALUES button towards the bottom middle of the LAYER PROPERTIES window to open the ADD VALUES window. From the list of values, scroll down until you find -10, and click on it. Hold the CTRL button down on your keyboard and select -20, -50, -100 and -150 as well. Now click OK.

You can now select colours for these contours in the LAYER PROPERTIES window. In the LAYER PROPERTIES window under SYMBOL, click on the box to the left of <ALL OTHER VALUES> to remove the tick inside it. This means that only the contours you have selected will be displayed. Now, double-click on the horizontal line next to each contour value and select a different shade of blue for each one, going from a light shade for the shallowest depth to a dark shade for the deepest depth. Once you have done this, click OK to close the LAYER PROPERTIES window.

Exercise One: Creating A Presence-Absence Raster Data Layer For A Species From Survey Data



At the end of this step, re-arrange the order of your data layers in the TABLE OF CONTENTS window so that appear in the order shown in the image below. This is done by clicking on a data layer's name and dragging it up or down to the desired new position.

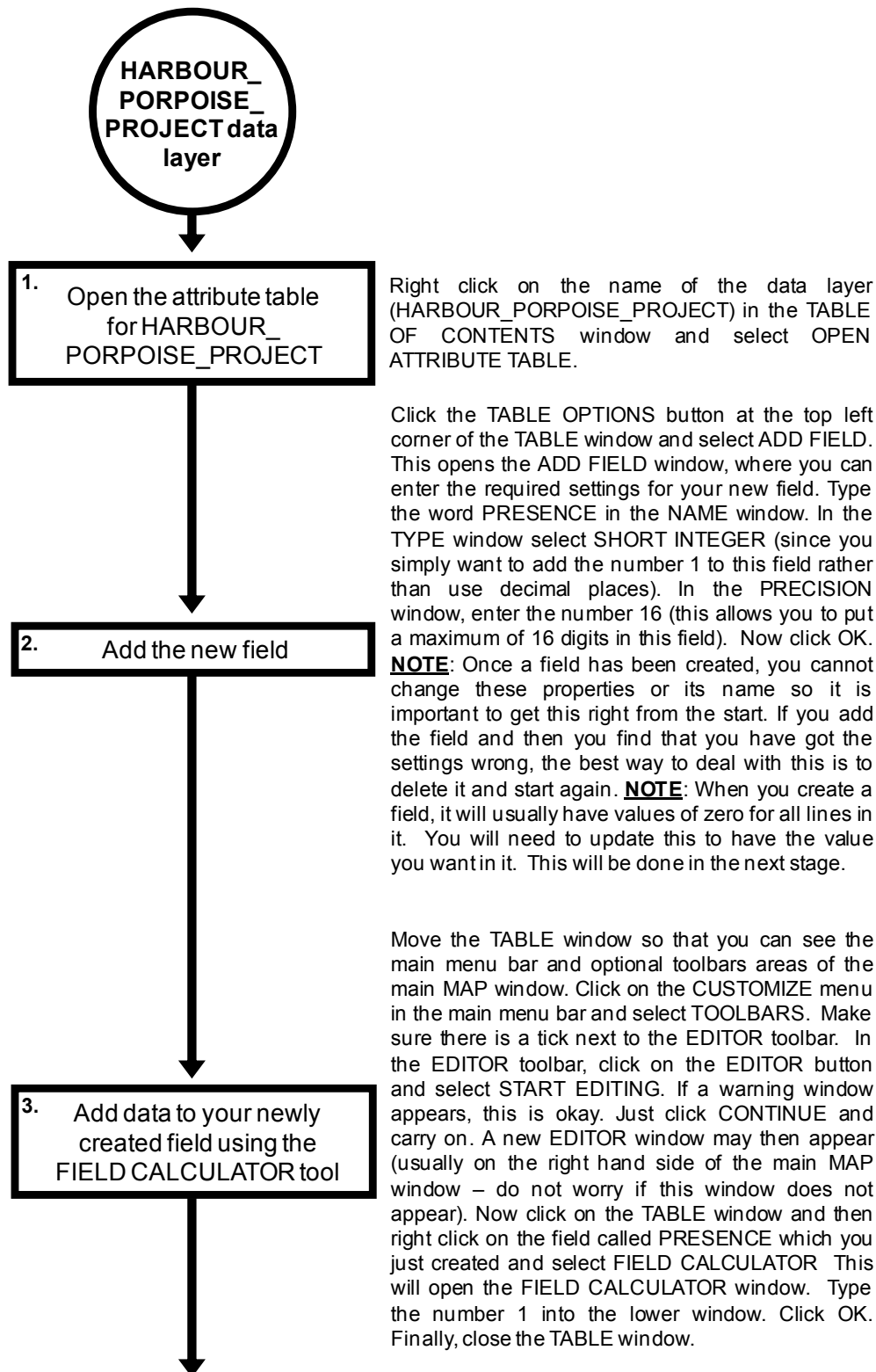


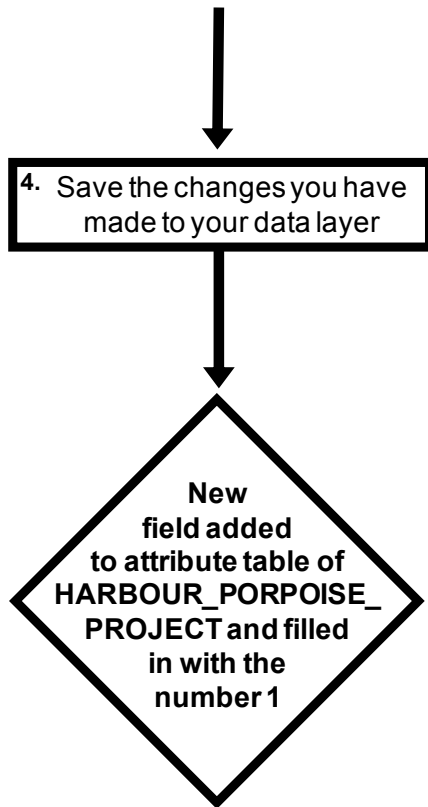
Next right-click on the name of the DEPTH_NORTH_SEA_PROJECT data layer in the TABLE OF CONTENTS window and select ZOOM TO LAYER. The contents of your MAP window should look like the image at the top of the next page.



STEP 3: ADD A NEW FIELD TO THE ATTRIBUTE TABLE OF THE SPECIES LOCALATIONAL RECORDS DATA LAYER AND FILL IT WITH THE NUMBER 1:

In order to make a presence-absence raster data layer, you first need to create a presence raster data layer for your species, this will be done in step four. However, before you can do this you need to have a field in the attribute table of your data layer of the species locations (i.e. HARBOUR_PORPOISE_PROJECT) which has a value of one in it for all records. This indicates species presence and is created in this step. The instruction set for this step is primarily based on one called *How to add a new field to an attribute table*, but it also includes steps from *How to use the field calculator tool to fill in values in a new field*. Generic versions of these instruction sets can be found in *An Introduction To Using GIS In Marine Biology*.





Now that you have finished editing the attribute table for the HARBOUR_PORPOISE_PROJECT data layer, you need to save the edits. To do this, in the EDITOR toolbar, click on the EDITOR button and select SAVE EDITS. Next, in the EDITOR toolbar, click on the EDITOR button again and select STOP EDITING. **NOTE:** If you were adding a lot of data, you should save the changes to the table as you go along.

NOTE: If the EDITOR window (sometimes also labelled as the CREATE FEATURES window) remains open once you have stopped editing the data layer, close it before you carry on with this exercise.

When you have completed this step, the attribute table for the HARBOUR_PORPOISE_PROJECT data layer should have changed from looking like the top half of the figure at the top the next page to looking like the bottom half, with a new field called PRESENCE added, which has a 1 in it for each line. If the attribute table is not open, right click on the name HARBOUR_PORPOISE_PROJECT in the TABLE OF CONTENTS window and select OPEN ATTRIBUTE TABLE.

Exercise One: Creating A Presence-Absence Raster Data Layer For A Species From Survey Data

FID	Shape *	Latitude	Longitude	Species	Number
0	Point	57.333167	-1.785333	Harbour porpoise	1
1	Point	57.145167	-2.061167	Harbour porpoise	2
2	Point	57.151667	-2.044333	Harbour porpoise	2
3	Point	57.196833	-1.989333	Harbour porpoise	1
4	Point	57.206167	-1.989333	Harbour porpoise	1
5	Point	57.251833	-1.921167	Harbour porpoise	1
6	Point	57.2684	-1.903	Harbour porpoise	1
7	Point	57.27835	-1.888383	Harbour porpoise	1
8	Point	57.28325	-1.88235	Harbour porpoise	1
9	Point	57.298917	-1.862983	Harbour porpoise	2
10	Point	57.315183	-1.842783	Harbour porpoise	5
11	Point	57.321067	-1.835633	Harbour porpoise	3
12	Point	57.35925	-1.788	Harbour porpoise	1
13	Point	57.391533	-1.745483	Harbour porpoise	1
14	Point	57.3975	-1.737767	Harbour porpoise	1
15	Point	57.427183	-1.698833	Harbour porpoise	1
16	Point	57.46375	-1.6494	Harbour porpoise	2
17	Point	57.472067	-1.638033	Harbour porpoise	2
18	Point	57.494533	-1.630833	Harbour porpoise	1
19	Point	57.4867	-1.630717	Harbour porpoise	1
20	Point	57.537567	-1.625433	Harbour porpoise	2

↓

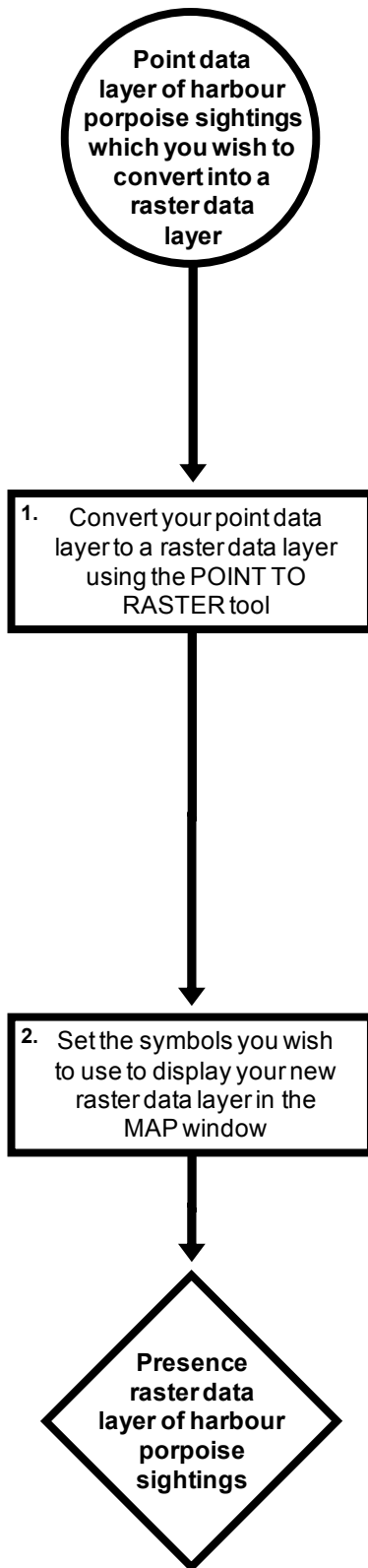
FID	Shape *	Latitude	Longitude	Species	Number	PRESENCE
0	Point	57.333167	-1.785333	Harbour porpoise	1	1
1	Point	57.145167	-2.061167	Harbour porpoise	2	1
2	Point	57.151667	-2.044333	Harbour porpoise	2	1
3	Point	57.196833	-1.989333	Harbour porpoise	1	1
4	Point	57.206167	-1.989333	Harbour porpoise	1	1
5	Point	57.251833	-1.921167	Harbour porpoise	1	1
6	Point	57.2684	-1.903	Harbour porpoise	1	1
7	Point	57.27835	-1.888383	Harbour porpoise	1	1
8	Point	57.28325	-1.88235	Harbour porpoise	1	1
9	Point	57.298917	-1.862983	Harbour porpoise	2	1
10	Point	57.315183	-1.842783	Harbour porpoise	5	1
11	Point	57.321067	-1.835633	Harbour porpoise	3	1
12	Point	57.35925	-1.788	Harbour porpoise	1	1
13	Point	57.391533	-1.745483	Harbour porpoise	1	1
14	Point	57.3975	-1.737767	Harbour porpoise	1	1
15	Point	57.427183	-1.698833	Harbour porpoise	1	1
16	Point	57.46375	-1.6494	Harbour porpoise	2	1
17	Point	57.472067	-1.638033	Harbour porpoise	2	1
18	Point	57.494533	-1.630833	Harbour porpoise	1	1
19	Point	57.4867	-1.630717	Harbour porpoise	1	1
20	Point	57.537567	-1.625433	Harbour porpoise	2	1

Now close the attribute table for the HARBOUR_PORPOISE_PROJECT data layer.

STEP 4: CONVERT HARBOUR PORPOISE DATA LAYER TO A RASTER DATA LAYER BASED ON THE NEW FIELD CALLED 'PRESENCE':

In this step you will create a raster data layer where the cell values will indicate whether harbour porpoise were recorded as present during the study. It will use the values in the PRESENCE field of the attribute table of the HARBOUR_PORPOISE_PROJECT data layer, created in step three, to calculate the value for each cell, and when more than one harbour porpoise record falls within the same cell, it will calculate an average value for that cell. Since all the harbour porpoise records in the HARBOUR_PORPOISE_PROJECT data layer have a value of one for the presence field, this will result in a cell value of one regardless of how many sightings were recorded in an individual cell. A cell size of 1km by 1km (or 1,000m by 1,000m) will be used for the raster data layer's resolution. This cell size was selected as it provides a raster data layer which, based on existing knowledge of this species in this study area, is appropriate to both the distribution of harbour porpoises in this region and the spatial variation in environmental variables. The extent will be set to cover an area which includes the main survey routes between Aberdeen (in the south) and the islands of Orkney and Shetland in the north. This will be the area covered by the final species distribution model. In terms of a projection/coordinate system, it will use the same one as the data frame. This is because it allows you to create a raster data layer which is measured in real world units (in this case metres) and is appropriate to the size and position of the study area. In the raster data layer created in this step, which will be called HP_RASTER, any cells where harbour porpoises were recorded will have a value of one, and all other cells will be classified as NO DATA, and will not be displayed.

The instructions for this step (which start on the next page) are based on instruction sets called *How to create a new raster data layer from an existing point data layer* and *How to change the way a raster data layer is displayed*. Generic versions of these instruction sets can be found in *An Introduction To Using GIS In Marine Biology*. The generic instruction set for *How To create a new raster data layer from an existing point data layer* includes advice for selecting appropriate cell sizes, extents and projection/coordinate systems for a specific data set and study area. However, for this exercise, you will use the ones detailed above.



In the toolbox window, select CONVERSION TOOLS > TO RASTER > POINT TO RASTER. In the POINT TO RASTER window, select the point data layer called HARBOUR_PORPOISE_PROJECT in the INPUT FEATURES window using the drop down menu. Select the field called PRESENCE using the drop down menu in the VALUE FIELD window. Type C:\GIS_EXERCISES_3\HP_RASTER into the OUTPUT RASTER DATASET window. In the CELL ASSIGNMENT TYPE (OPTIONAL) window select MEAN. Type the number 1000 into the CELLSIZE (OPTIONAL) window.

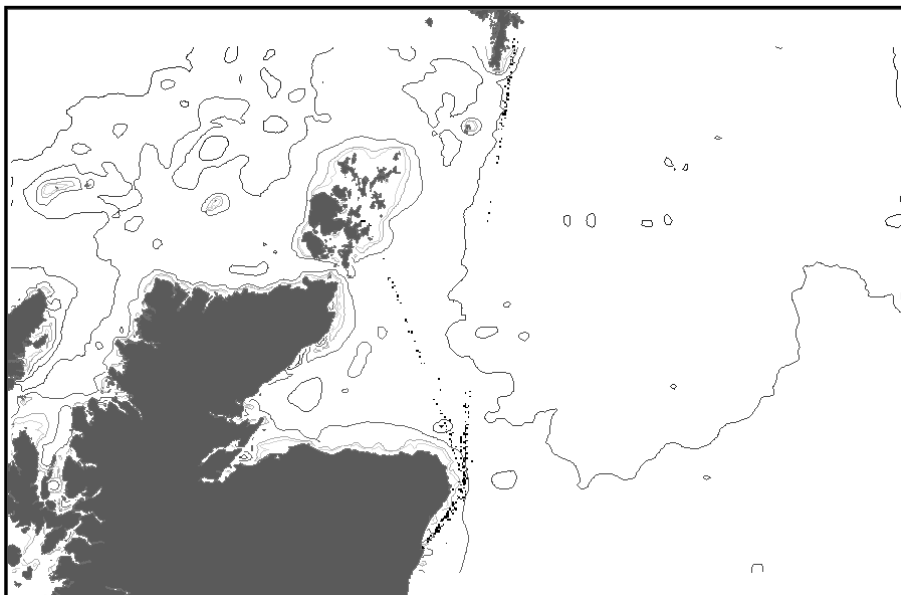
Next, click on the ENVIRONMENTS button at the bottom of the window. In the ENVIRONMENT SETTINGS window, click on PROCESSING EXTENT. In the EXTENT section of the window that will appear select AS SPECIFIED BELOW from the drop down menu, and type in the following coordinates: TOP: 396000, LEFT: -130000, RIGHT: 50000 and BOTTOM: 56000. Now click OK to close the ENVIRONMENTS SETTINGS window. Then click OK at the bottom of the POINT TO RASTER window.

Right click on the name of your newly created raster data layer (HP_RASTER) in the TABLE OF CONTENTS window and select PROPERTIES. **NOTE:** It may have been added below other data layers you already have in your data frame. Next, click on the SYMBOLOGY tab of the LAYER PROPERTIES window. In the left hand portion of the window, select UNIQUE VALUES. When it asks 'UNIQUE VALUES DO NOT EXIST. DO YOU WANT TO COMPUTE UNIQUE VALUES?' click YES. Next, click on the ADD ALL VALUES button. Double click on the coloured rectangle beside the number 1, and select black for the colour. Finally, click the OK button. All the cells in the HP_RASTER data layer where harbour porpoise were present will now be coloured black. All other cells will be classified as no data and will not be shown.

Exercise One: Creating A Presence-Absence Raster Data Layer For A Species From Survey Data

Once you have completed this step, you can turn off both the HARBOUR_PORPOISE_PROJECT and SURVEY_TRACKS_NORTH_SEA_PROJECT data layers so that you can see your presence raster data layer properly. This is done by clicking on the small box with a tick in it next to their names in the TABLE OF CONTENTS window so that the tick disappears.

The contents of your MAP window should now look like this:



NOTE: In this step, two major decisions have been taken that can help determine if your SDM project will succeed. These are to set the extent of your study area and to decide on the resolution (as determined by the cell size) at which you will model the distribution of your target species. Neither of these decisions should be taken lightly and if you are doing your own SDM project, you will need to ensure that both the extent and the resolution you select in this step are appropriate for the aims of your study, your specific study area and the species which you are aiming to model. If you do not put enough thought into selecting them at this stage of your SDM project, it is likely that you will have to repeat it again at a later point in your project if you find out that you have selected the wrong values.