**Systematic Conservation Planning Workshop Exercise: Understanding and Assembling Model Input**

**Introduction**

In this exercise, we will demonstrate the many data acquisition and pre-processing that are generally required for systematic conservation planning analysis. These exercises are designed to provide a guide on some of the most common steps that are performed before more advanced analyses.

The data used in this activity are for educational purposes only, and in some cases may lack the quality required to produce accurate and precise results.

**Part 1: Acquiring Occurrence locations**

There are many online and free sources to find spatially reference species occurrences. These range from local to global and can be a valuable source of data. Here, we will demonstrate how to download occurrences locations for *Grus canadensis* from the Global Biodiversity Information Facility (GBIF)

Go to <http://www.gbif.org/>

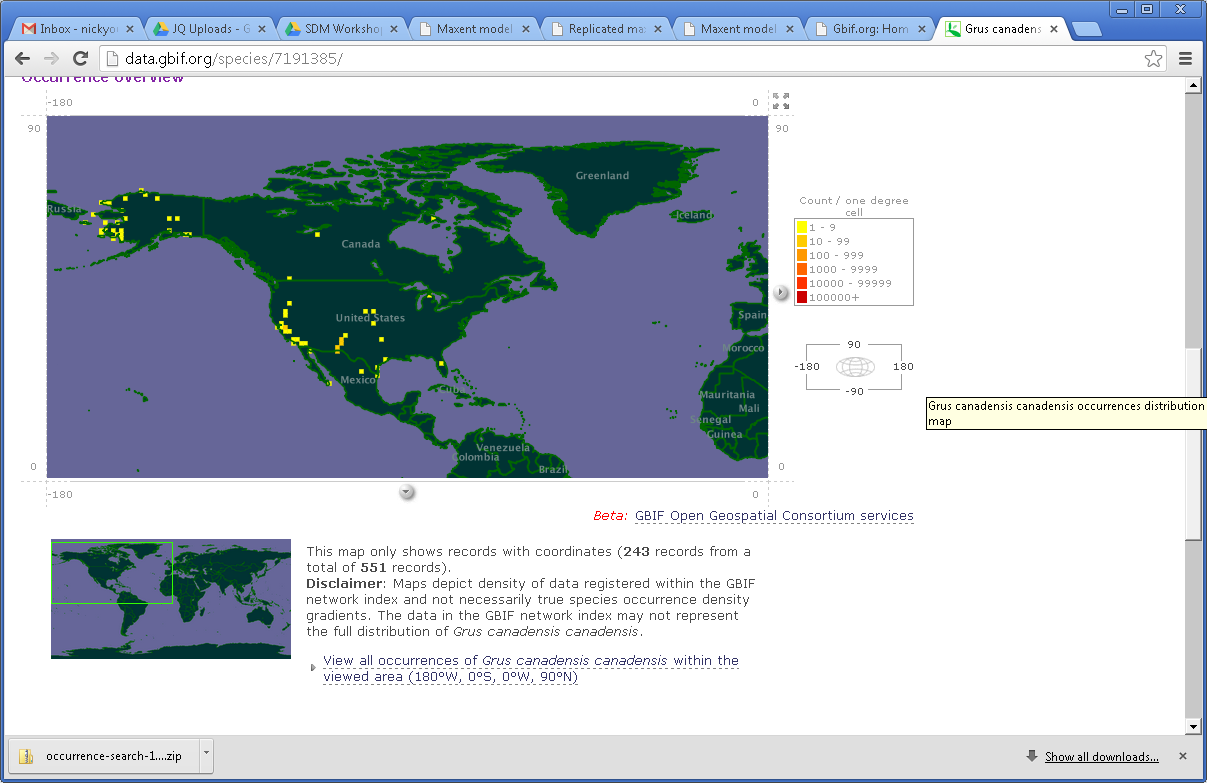
On the right side of the page you should see a big “**access data portal**” button. Click this link.



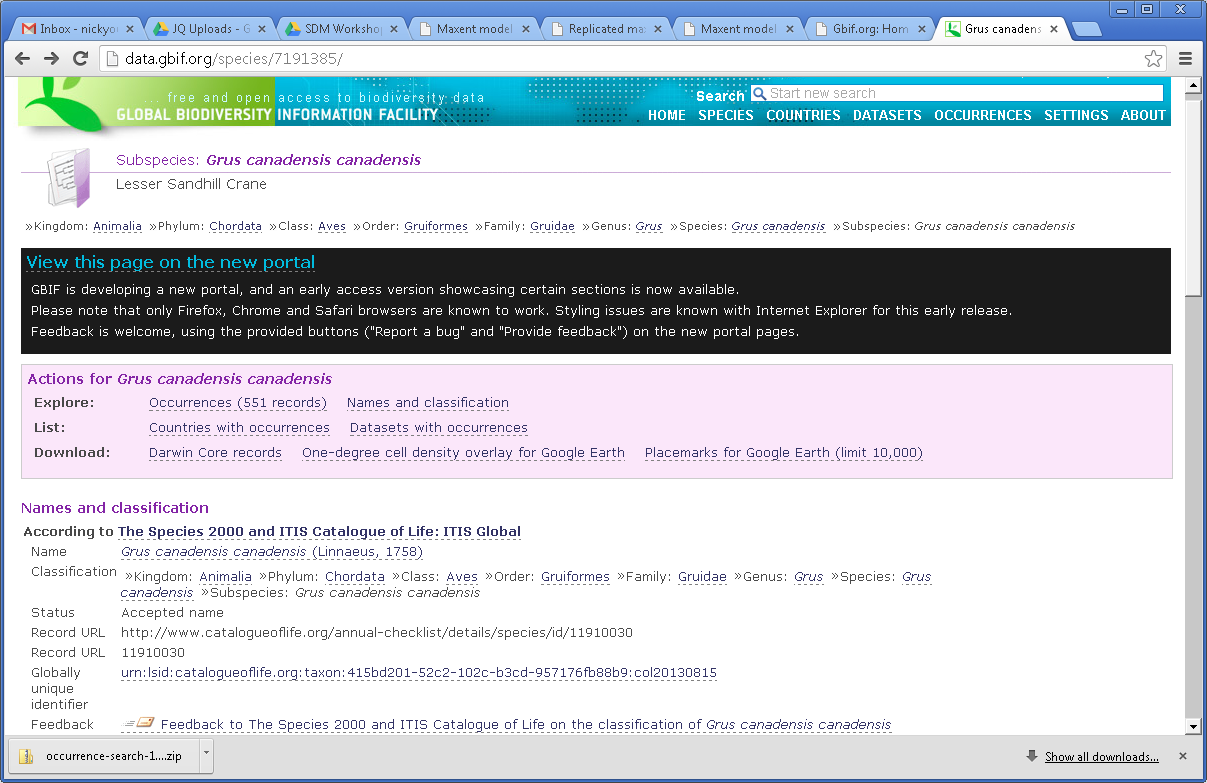
Search for *Grus canadensis* in the search box. The next page is a terms of use page, **Accept** the terms.

You will see a list of a variety of results. For this purpose, let’s say we are only interested in the lesser Sandhill crane (*Grus canadensis canadensis*). Select this subspecies from the list.

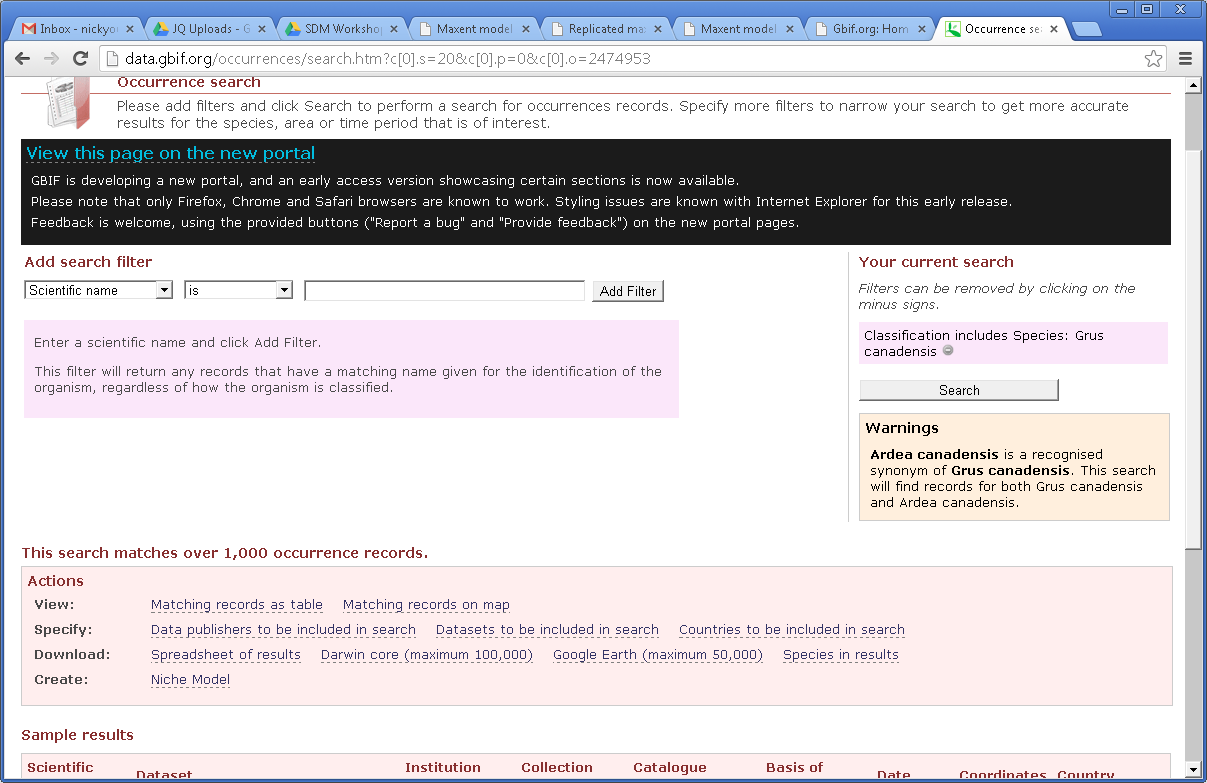
This will bring you to a page with some taxonomic and distribution information. Scroll down to look at the global map of occurrences for a quick inspection.



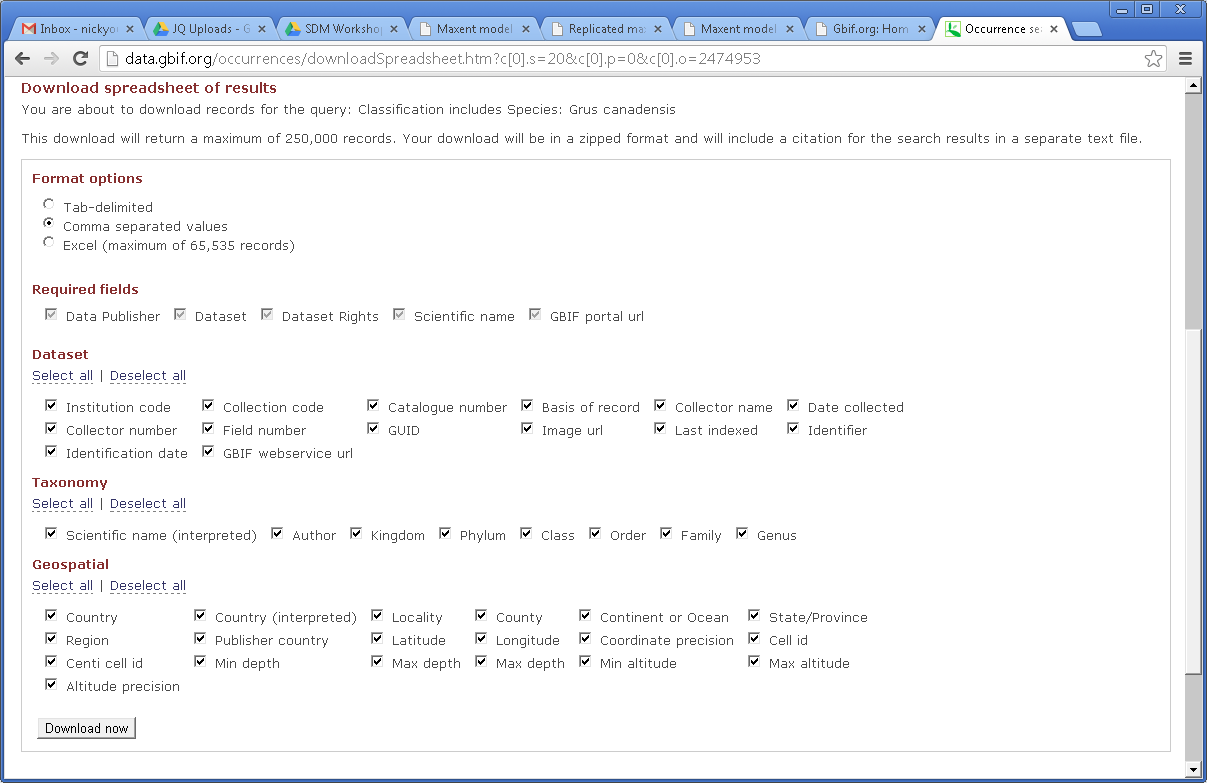
At the top of the page there is a box titled “**Actions for *Grus canadensis canadensis***”. Click on the Occurrences link under **Explore:**



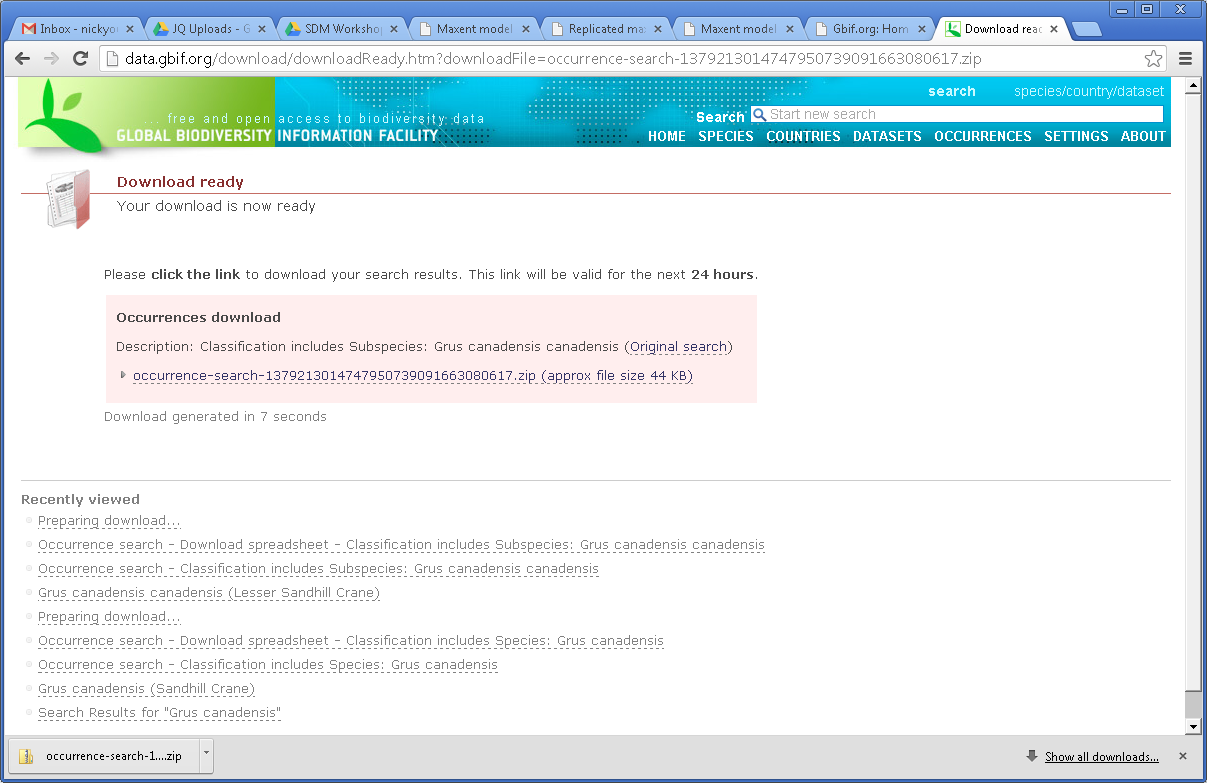
On this next page, we could add some filters to only include a certain institution for example. We are interested in the full dataset, so we won’t provide any filters. Scroll down to the **Actions** section and select **Spreadsheet of results** under **Download:**



The next page provides a variety of options for the downloaded data format. From a selection of the specific attributes to include to the type of file the data will be delivered as. Choose **Comma Separated Value** format and leave the rest of the defaults. Click the **Download Now** button at the bottom of the page

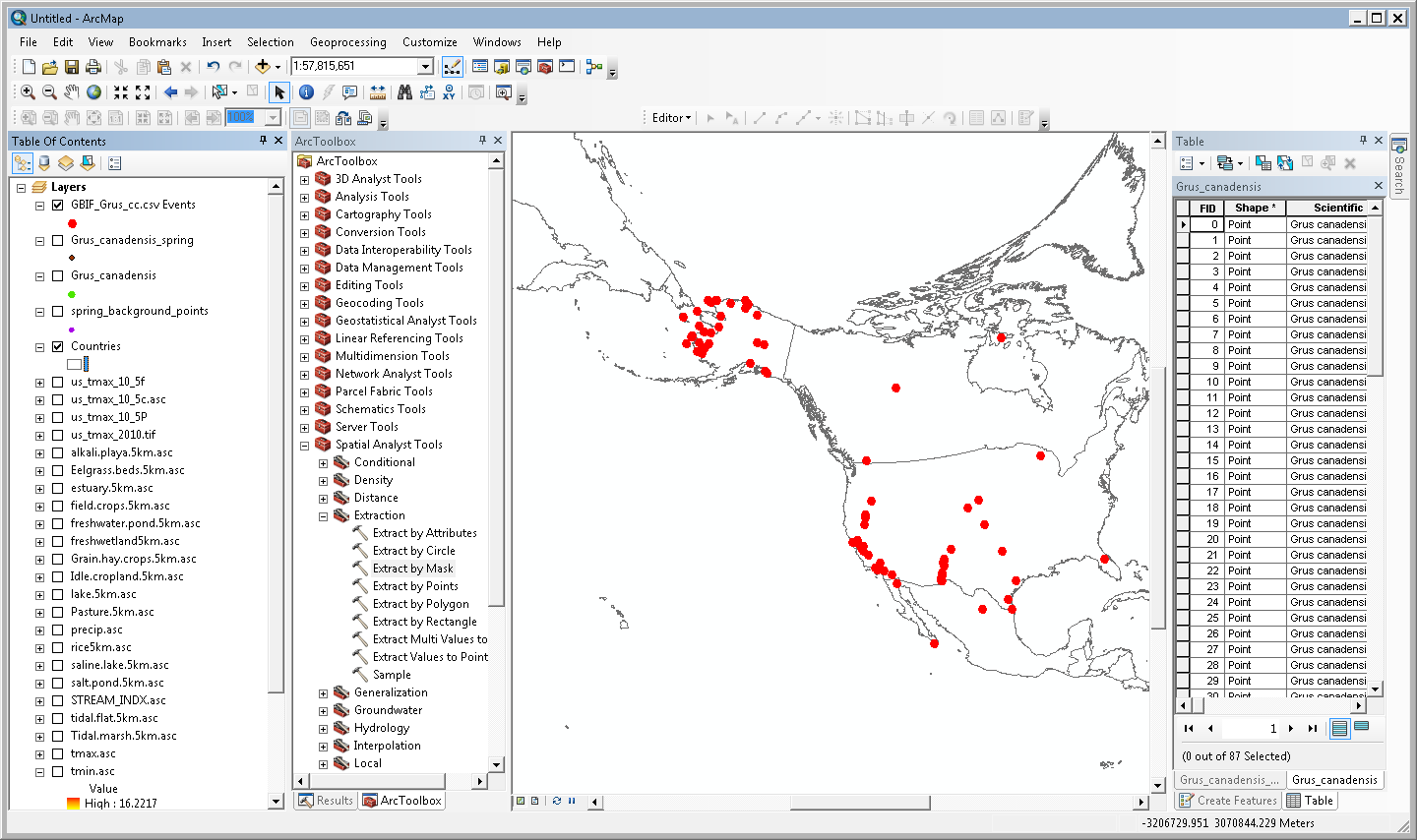


You’ll be brought to a download page that will let you know when your request is available for download.

****

Save the file in the **Workshop\_StartData**. The data comes as a zipped file that will need to be extracted (unzipped). Extract the files to the same location and **open** the .csv file for inspection.

* **What is the coordinate reference system of the data? How would you find this out?**
* **How many records don’t have any coordinates?**
* **Are there any duplicate records?**
* **How many records are unique?**
* **What are some field that might help provide a measure of data quality/accuracy**
* **What is the accuracy of a location with a latitude of 31.9 vs. 29.73457? How might this impact any analysis performed with these data?**
* **Below is the map of the occurrences. Are there any points the look suspect to further evaluation?**

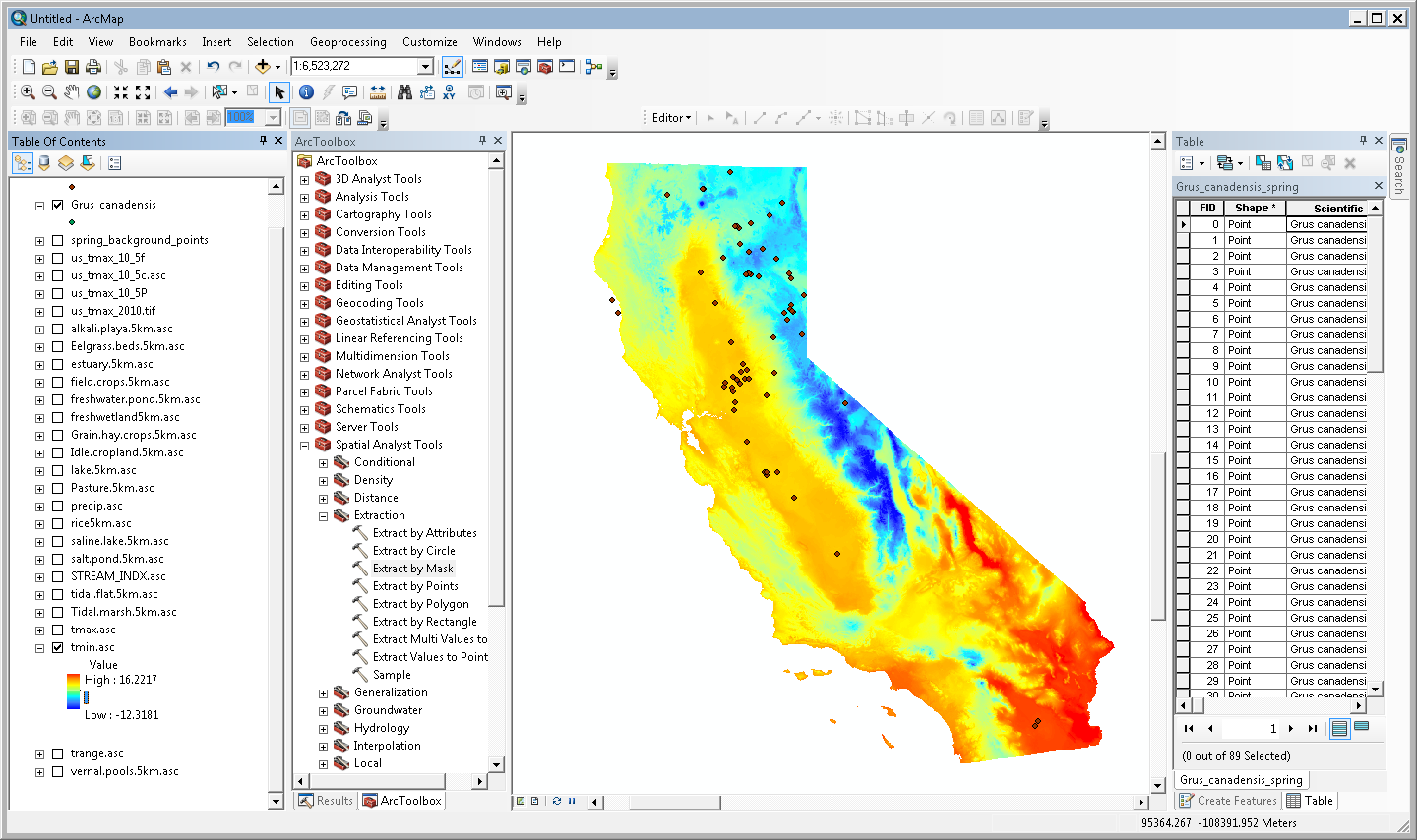
****

**Part 2: Occurrence data exploration.**

A critical step in any modeling exercise should be to explore the data that will be used for modeling. We are only going to touch on this subject but encourage rigorous evaluation of the data before proceeding to model generation.

First, open a new ArcMap document.

Load the **Grus\_canadensis\_spring.shp** file located in the **Workshop\_SartData** folder that consists of occurrence points for Sandhill crane in the spring months. To help with a frame of reference, load one of the environmental data layers in the **EnvironmentalData** folder



* **These data came from eBird which comprises of data contributed by volunteer and professional**
  + **What might be some concerns with this data set?**
  + **If you were handed this data set, what would be some questions you would ask to better understand the assumptions and limitations of the data?**
* **Are there any points that look suspicious?**
* **How many occurrences are there?**
* **Are there any duplicate locations? How might this impact a species distribution model? How might two occurrences that have the same location have different environmental data?**

**Part 3: Attributing environmental data to background locations.**

Now that you have examined the sample file, you need to make sure that you have the background data prepared for Maxent. We will discuss this subject more later in the workshop.

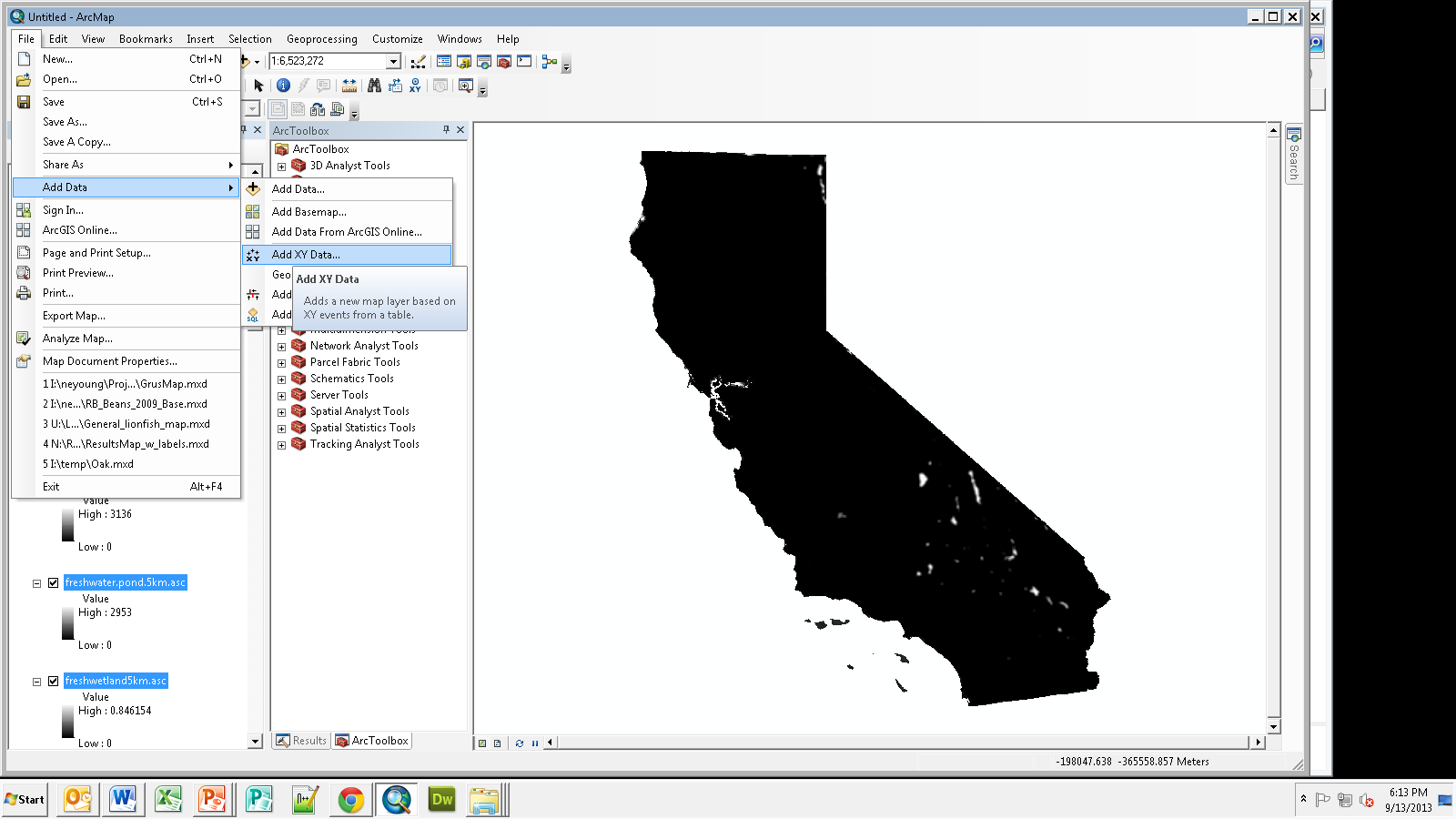
**Step 1: Loading occurrences from a spreadsheet**

First, open a new ArcMap document.

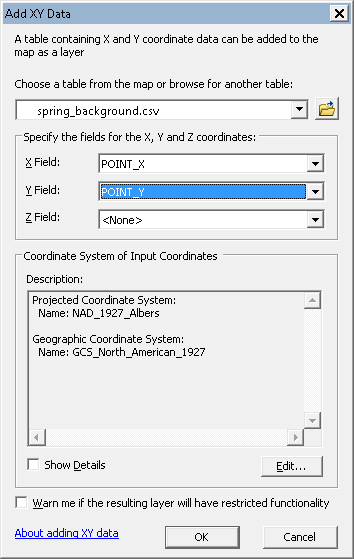
Add all the ascii layer in the **EnvironmentalData** folder. To do this, select the add data button and navigate to the **EnvironmentalData** folder within the **Workshop\_StartData** folder. Add all the rasters to the map (you can add them all at the same time by selecting the first raster, holding the shift key and then click the last layer).

In order to perform the analysis to attribute the background data, you will need to convert the background file into a shapefile (.shp)

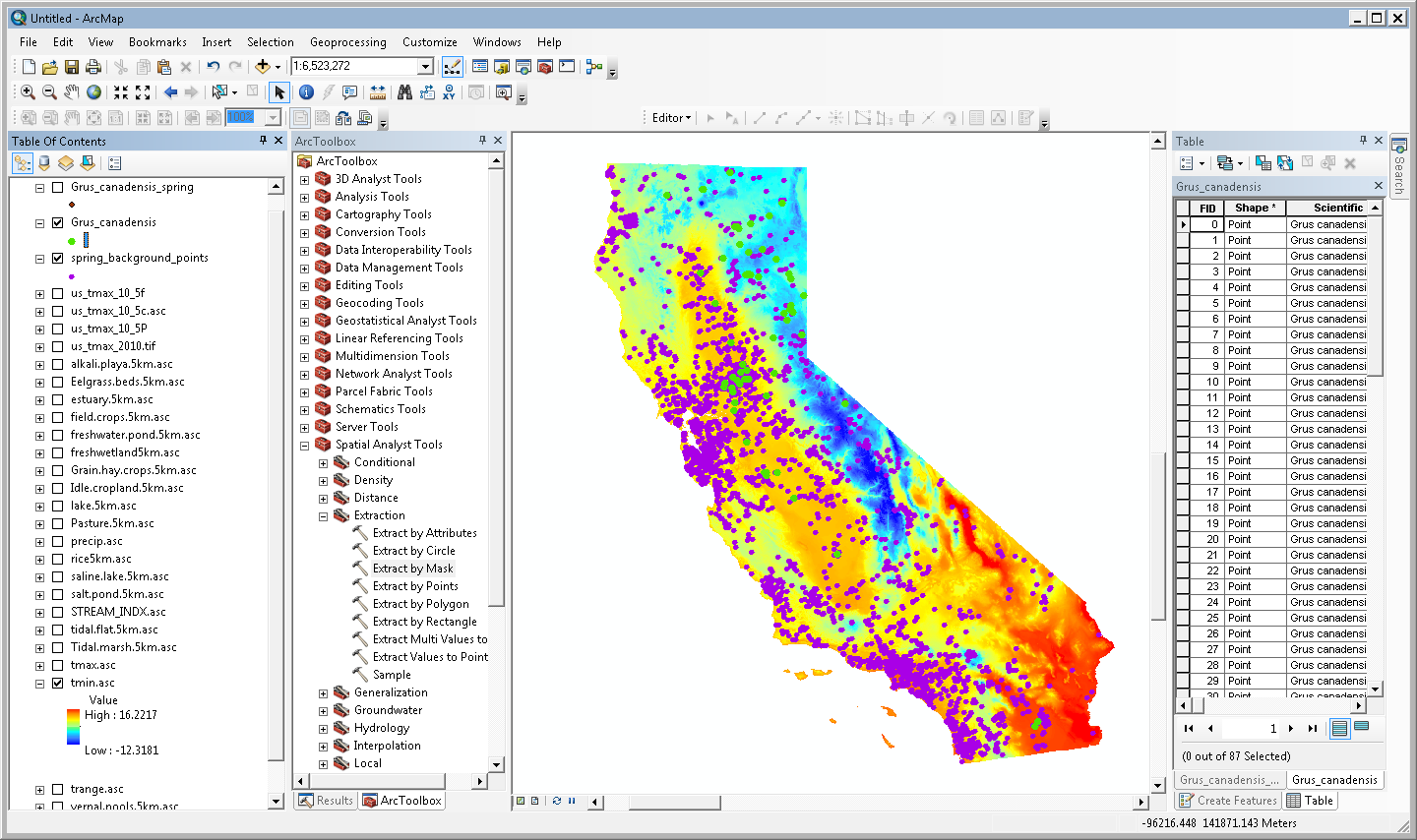
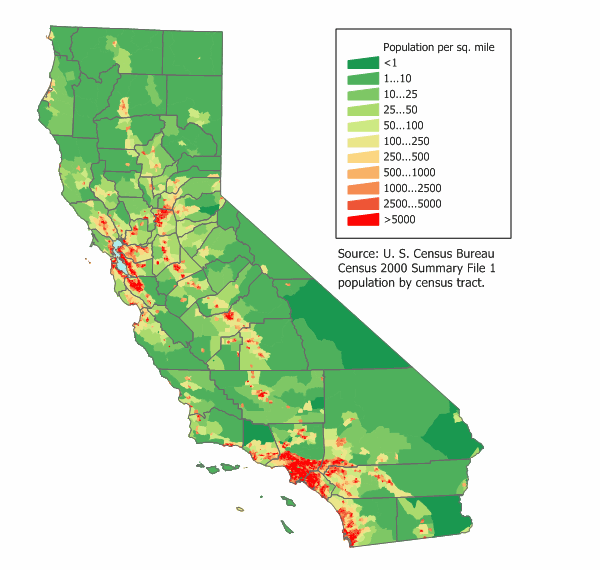
In the same ArcMap, select File🡪Add Data🡪Add XY Data…



A new window will appear. Navigate to the **spring\_background\_start.csv** file. In the window you will need to make sure that the **X field** and the **Y field** are set to the correct columns in the csv file. Click **OK**.



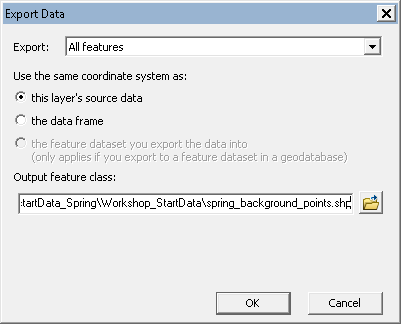
The background points should now be displayed on the map.

* **What do you notice about the distribution of the background points?**
  + **Are they a random sample of the environment?**
  + **Where are they concentrated?**
  + **How is the distribution of the background points similar to the occurrence points?**
  + **These points represent all eBrid observations for all other species during the spring months. Why would we want to use these points as our background sample?**

**Step 2: Creating a shapefile**

Next we will create a shapefile of these locations. Right-click the layer you just added called spring\_background.csv Events Select Data🡪Export data… A new window will appear. Click the folder icon for the **Output feature class** field to set the location and name of the shapefile. Save the file in the **Workshop\_StartData** folder and name it Convert background file to point layer. Click **OK**.



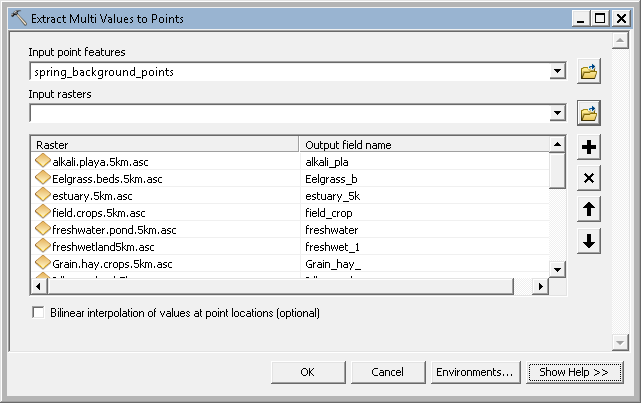
Once it has finished, it will ask you if you would like to add the layer to the map, select YES. You can remove the other Events layer at this point

**Step 3: Attributing the background points**

Now that we have both the background points and the environmental data loaded we will want to perform the attributing analysis using the **Extract Multi Values to Points** tool.

Search for “*extract multi*” in the Search window (this window may not be visible. If not, click on the icon in the toolbar or simply press **ctrl+f**). You should see the **Extract Multi Values to Points tool** as the first option listed. Select this item to open the tool.

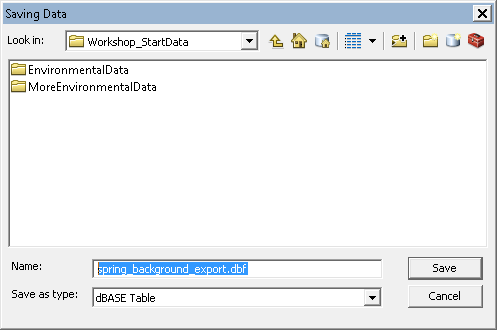
For the **Input point features**, choose the *spring\_background\_points.shp* file. For the **Input rasters** field add all the environmental raters you added earlier. Click **OK**.



This may take some time to complete.

**Step 4: Exporting a table and creating a comma separated value file**

Once it has finished, open the attribute table for the *spring\_background\_points.shp* file. Export the table by selecting the table options icon🡪 Export…Save the file in the **Workshop\_StartData** folder and name it “*spring\_background\_export.dbf*”. Make sure to choose **dBASE** file as the type



You will not need to add the table to the Map.

Next, navigate to where you saved the table file and open the file in excel. Remove any extra fields that may be present. The Maxent modeling software requires files to be in the comma separated value (.cvs) format. Select File🡪Save as… Save the file in the **Workshop\_StartData** folder and name it “*spring\_background\_forMaxent.csv*” and make sure you choose Comma Separated Value (.csv) file format.

This file is now ready for Maxent modeling software.

* **Take a look at the column names in the background sample spreadsheet. Do the match the original names of the environmental rasters?**
* **Maxent requires the names of the columns in the background file to match the file names of the environmental rasters. What would we need to do to make sure we don’t hit any errors when running Maxent**

**Part 4: Modifying Environmental Layers to be the Same Extent (geographic bounds and cell size) Using ArcGIS**

**Background:**

This part of the tutorial walks you through the steps required to modify an environmental raster grid in ArcMap so that all your spatial environmental data (i.e., independent or predictor variables) have the same extent and resolution (same geographic bounds and cell size). MaxEnt requires all the environmental layers be in raster format and have the exact same cell size, extent, and projection system (e.g., geographic or UTM) in order to execute a model.

We will be using just one layer to demonstrate this example

**Step 1: Loading layers into ArcMap and opening the Extract by Mask tool**

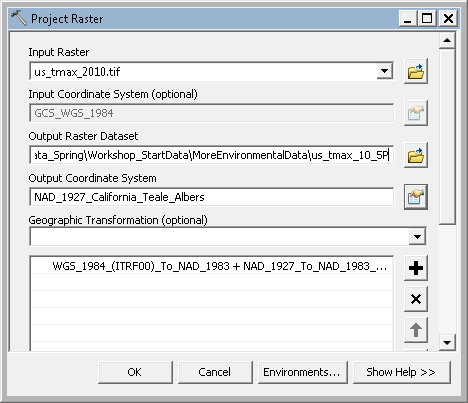
Open a new map in ArcMap. Click on the add data icon  and navigate to the MoreEnvironmentalData folder. Add the *us\_tmax\_2010.05.tiff* file. As a comparison, load one of the layers in the EnvironmentalData folder

* **How do the file formats between the two layers compare?**
* **What is the extent of the us\_tmax\_2010.05.tiff and how does this compare to the other rasters?**
* **Check the grain (cell size) of the two layers – any differences?**
* **What about the coordinate reference system between the two layers?**

Obviously there are some difference that we will need to correct if would like to include this raster in our modeling process.

**Step 2: Re-projecting the match coordinate reference systems**

First you will have to re-project the *us\_tmax\_2010.05.tiff* to match the rasters. To accomplish this, use the re-project tool in Data Management Tools🡪Projections and Transformations🡪Ratter🡪Project Raster. Open this tool. For the Input Raster, add the *us\_tmax\_2010.05.tiff* file. The Input Coordinate system should automatically be filled. For the Output Coordinate Dataset, Save the file in the MoreEnvironmentalData folder and name it **us\_tmax\_10\_5P**. For the Output Coordinate System, use the NAD\_1927\_California\_Teale\_Albers projection (Projected Coordinate System🡪State Systems). Once you have this filled out, click OK.



**Step 3: Converting a raster to ASCII**

Now we need to convert the file format from tiff to ASCII. We will accomplish this by using the Raster to ASCII too.

In the Toolbox window double click **Conversion Tools** then **From Raster,** andthen double click on the **Raster to ASCII** tool.

For the **Input Raster** click on the folder icon  and navigate to the *us\_tmax\_10\_5P* file in the **MoreEnvironmentalData** folder.

For the **Output ASCII** raster file, click on the folder icon  and navigate to the same **MoreEnvironmentalData** folder but name the file *us\_tmax\_10\_5c***.** Make sure to change the Save as type from .TXT to .ASC. Click **Save.**

Now that we have out new environmental raster in the same coordinate reference system we will resample and clip the layer to match those we already have.

**Step 4: Resampling and Clipping (extract by mask)**

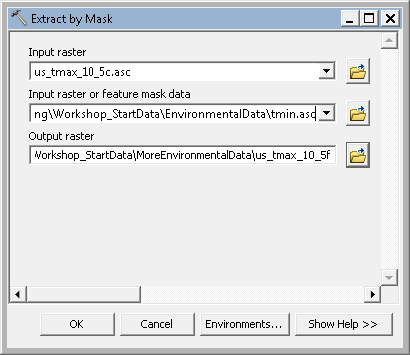
If your environmental layers happen to be larger than the area you are interested in modeling, this step allows you to clip them down to your area of interest and set all layers to have the same extent, cell size, and coordinate system (a requirement of MaxEnt). This step can also be used to set your extent, cell size and coordinate system if your environmental layers are already clipped.

To start, open the **Extract by Mask** tool (You can search for this tool or find it in the toolbox: Spatial analysts Tools🡪Extraction). A window should appear that looks like the one below for the **Extract by Mask** tool (location described above).

For the **Input raster** click on the folder icon  and browse to the **MoreEnvironmentalData** folder and select the *us\_tmax\_10\_5c.asc* raster.

For the **Input raster or feature mask data** field, use the  icon to browse to a raster or polygon that represents the spatial boundary you intend to model. For our example this can be any layer in the **EnvironmentalData** folder

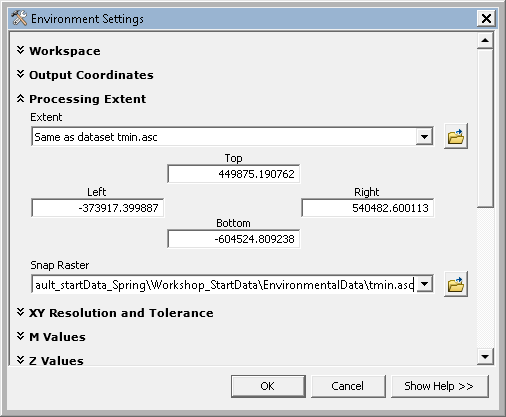
Save the **Output raster** as *us\_tmax\_10\_5f* under the folder **MoreEnvironmentalData**. Note, because Arc requires a ESRI GRID output for this tool, another step of converting the output to ASCII would be required before including in with the others for Maxent modeling.

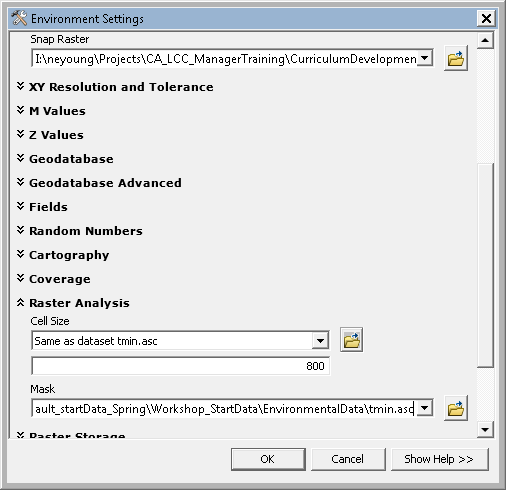


In order to set the cell size and extent, click on the **Environments…** button in the bottom right of the window\*. Extend the variables **Processing Extent**, and **Raster Analysis**.

For the **Processing Extent** and **Snap Raster** variables, browse to a raster you know that already has the extent you wish to match (such as the one you used as the mask).

For the field of **Cell Size** under the **Raster Analysis** variable**,** choose the same layer as you did for the processing extent. The true cell size value will appear in the window below the input field. It will reflect the units from which your coordinate system is based.





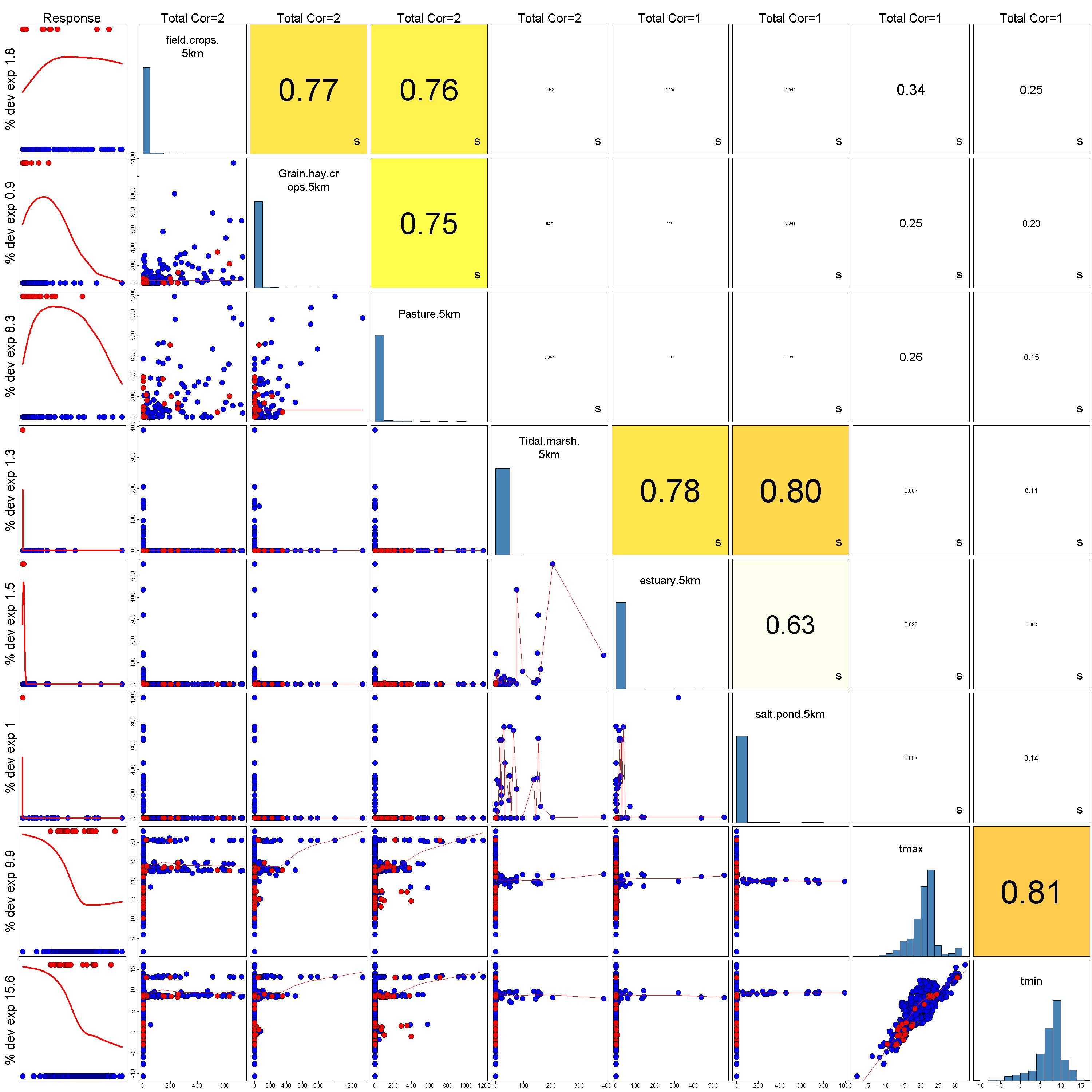
Click **OK**.

* **Once the tool has finished running, check the properties of the output:**
  + **Do the number of rows and columns match the other layers?**
  + **Is grain size the size the same as the other layers?**
  + **Coordinate Reference System**
* **What might some issues with using this method for resampling? Are we forcing a downscale or upscale of our new raster?**
* **Is there only one way to resample?**
* **What if this was a categorical variable? How would we want to modify our methods to make sure we are using the appropriate resampling methods?**

**Part 5: Evaluating and removing Environmental data**

In this section, you will examine the correlation structure of the environmental data in relation to the occurrence locations and background locations.

Collinearity is an important component to evaluate in the environmental layers. Below is a matrix of the environmental layers that we are considering for a spring distribution model for *Grus Canadensis* (Sandhill Crane) generated by another freely available tool, the software for assisted habitat modeling (SAHM). Note this only shows 8 of the 21 variables that we are considering for thee model. The remaining variables are not highly correlated.

****

As you can see, some of the variables have higher correlations than others. The general rule of thumb for SDM is to remove one of each pair of highly correlated variables (R2>0.70). When choosing which variables to remove, always try to keep the variables that you know have ecological relevance to the distribution of the species for the system of interest. Depending what is known about the species and on your knowledge of the species, this may be easy or difficult and there might not be enough information to know which environmental variable may be more important. In this case you can remove the variable that is more easily interpretable, the one that explains the most deviance, but one should be removed before continuing to the model.

**Spend some time deciding what variables you would remove and why. Write these down, you will use them later.**