Correlation between the Incidences of Invasive European Starlings and Human Populations in Florida

Blake Chapman

(Cabe, 1993)

I.   Introduction

Over 200 million European starlings inhabiting North America consist of many generations descended from about 16 known breeding pairs. These were the birds who successfully established themselves after an initial release of 100 individuals in New York across 1890 and 1891 (Linz et al., 2007). Starlings first reached Florida around 1920 and likely thrive due to their ability to live in a variety of habitats including agricultural lands and urban areas ("Nonnatives"). I wanted to investigate if European starling observations tended to appear more within cities and I used the Pearson Correlation Coefficient to examine if starlings occurred in the same patterns as human population densities for the year 2012.

II.   Data Collection and Preprocessing

Data was gathered from three sources. Point data on starling observations was obtained from the Early Detection and Distribution Mapping System (EDDMapS) through Distribution Maps tab (Figure 1). Information on birds is located underneath the Wildlife category. The Southwest Florida Water Management District (SWFWMD) provided a polygon shapefile of Florida from the Southwest Florida Water Management District's Geospatial Open Data Site subsection of their webpage (Figure 2). U.S. Census data on population densities of people for the year 2012 was downloaded from ArcGIS online (see ArcGIS Data Analysis for figures showing how to download the file).

Figure 1. EDDMapS page displaying where to download the starling point data
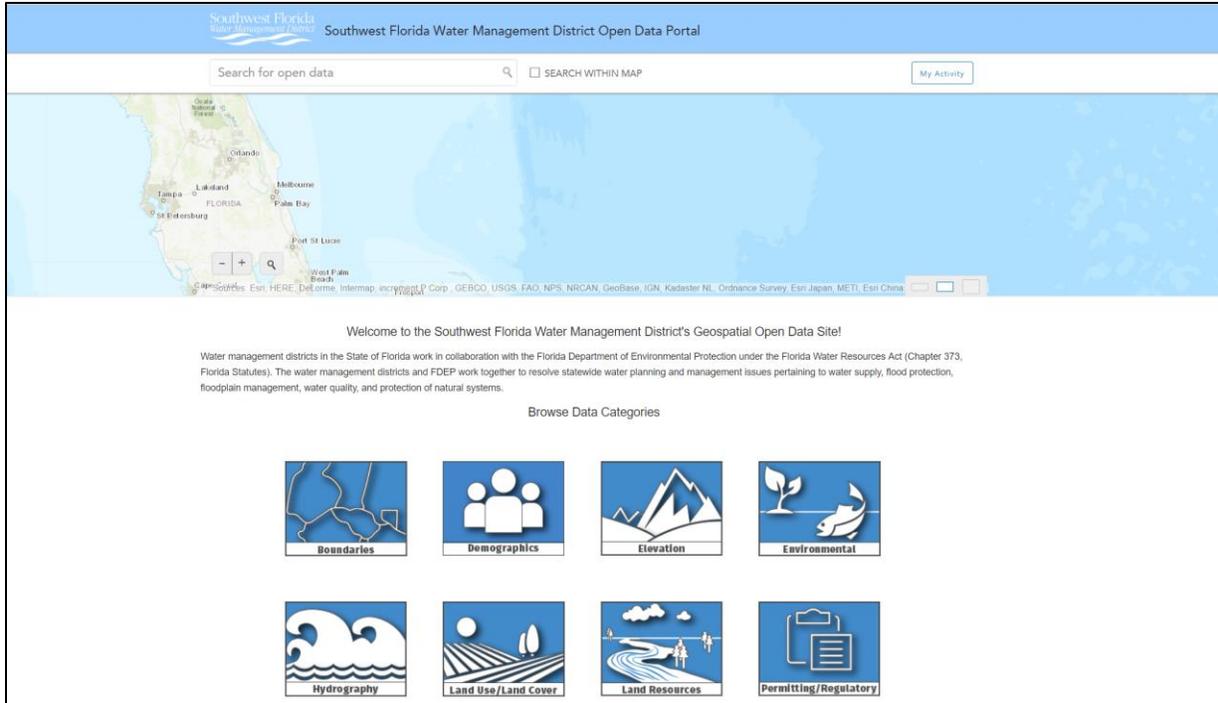
Figure 2. The Florida polygon shapefile can be found underneath the boundaries category from the SWFWMD

Preprocessing of the starling data occurred in a series of steps:

1. Add the starling point shapefile to ArcMap
2. Open the attribute table then and select and extract all the European starlings observed in 2012
3. Remove all the data from regions outside of Florida with the Clip (Analysis) tool
4. Deleted data listed to have originated from the US Army Corps of Engineers Ombil Database results due to lack of reported collection numbers. This leaves plenty of points to work with and does not affect the overall appearance of the map distribution.
5. Replaced very large and obviously incorrect values (nearly 19,000,000 starlings! [Figure 3a]) with the true collection values provided in separate reports (Figure 3b) for each point on the EDDMapS site under List. Points were identified using the date, observer, county, and coordinate provided in the attribute table.

IDENTIFICA	IDENTIFI_1	REVIEWED	REVIEWER	DATE_REVIE	NUMBER_COL
Credible	<Null>	Verified		<Null>	220
Credible	<Null>	Verified		<Null>	400
Credible	<Null>	Verified		<Null>	300
Credible	<Null>	Verified		<Null>	19050314
Credible	<Null>	Verified		<Null>	750
Credible	<Null>	Verified		<Null>	700
Credible	<Null>	Verified		<Null>	19070212
Credible	<Null>	Verified		<Null>	80
Credible	<Null>	Verified		<Null>	290
Credible	<Null>	Verified		<Null>	7
Credible	<Null>	Verified		<Null>	66
Credible	<Null>	Verified		<Null>	120
Credible	<Null>	Verified		<Null>	1
Credible	<Null>	Verified		<Null>	97
Credible	<Null>	Verified		<Null>	64

a)

european starling
Sturnus vulgaris Linnaeus, 1758

Species Information

States   Counties   Points   List

Revisit | Download Record (pdf)

Record ID   5155986
Location Information
Location   Charlotte County, Florida, United States
Coordinates   26.97482, -82.08739
Infestation Information
Population Size   1
Infestation Status   Positive
Reporter Information
Reporter   FWCC Exotic Species Database , Florida Fish and Wildlife Conservation Commission
Observation Date   April 12, 2017
Date Entered   August 23, 2017
Fwc Observation Id   279831
Source Type   Bulk Data
Species Information
Original Reported Subject   Sturnus vulgaris
Verification and Review
Verification Method   Expertise
Identification Credibility   Credible
Reviewed   Verified
Survey Information
Visit Type   Observation
Surveyor   Michelle Unknown
Other

b)

Figure 3. a) shows a couple of starling observations recorded at approximately 19,000,000 birds. These were corrected with reports that indicated population sizes of significantly smaller groups or individual animals (sample in b).

6. A few of the points had blanks underneath the number collected column. Most were corrected with the reports but those that could not were given assumed to be one as there must be at least a single individual to possess a reportable observation.
7. Created a new column with a short integer field and transferred the values from the NUMBER_COL column (string field) to obtain quantifiable numbers.

III.   ArcGIS Data Analysis

The investigation of starling-human population correlation began by adding the clipped and preprocessed starling 2012 and the Florida state polygon shapefile to a blank ArcMap (Figure 4).
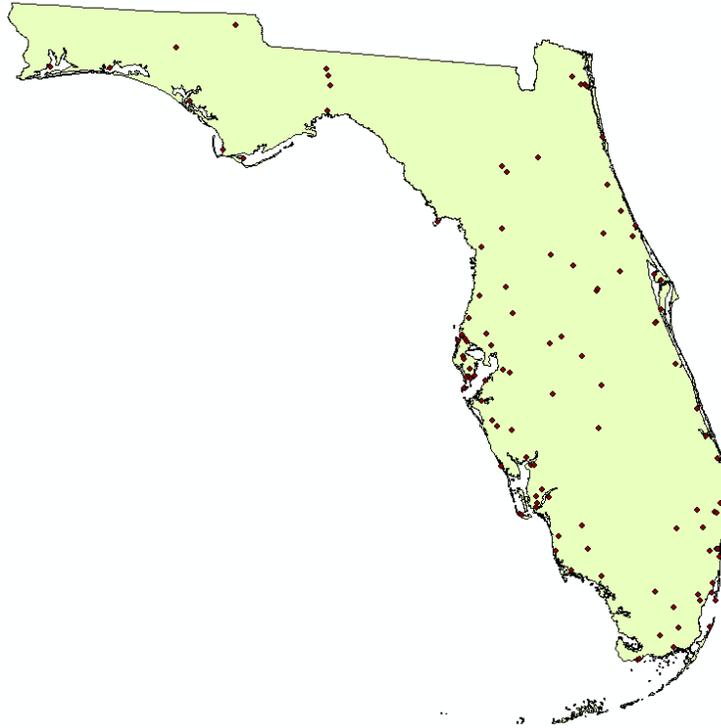
Figure 4. Starling observations are represented by the point on the map

Next, the Inverse Distance Weighting (IDW; Spatial Analyst) tool was utilized to interpolate occurrences of starlings between the data points (Figure 5a). After selecting the starling file, the program automatically adds the rest. Leaving the defaults including the arbitrarily chosen number of points, a mask was set to the Florida polygon under Environments (Figure 5b).
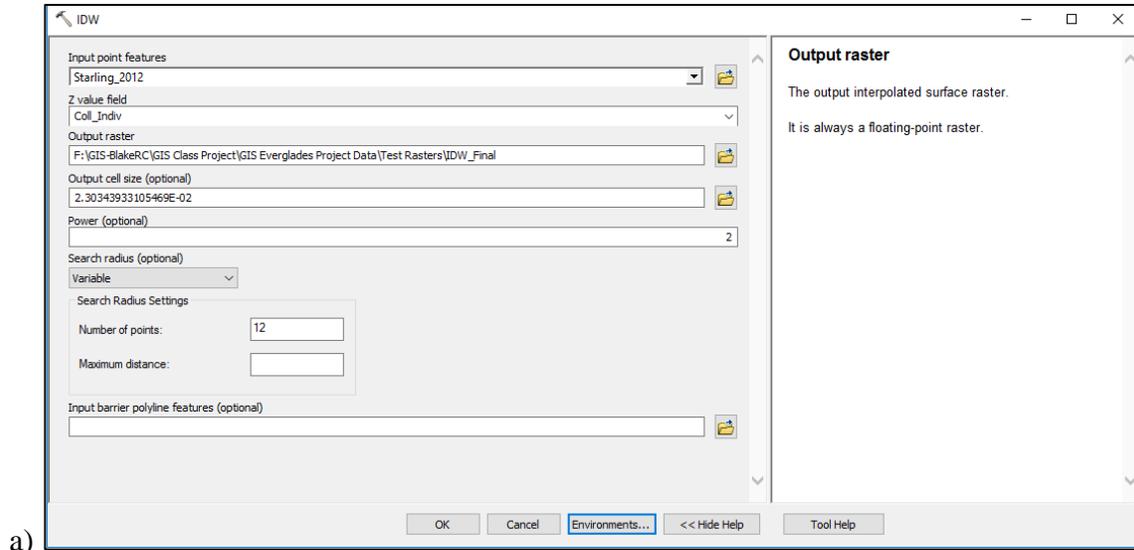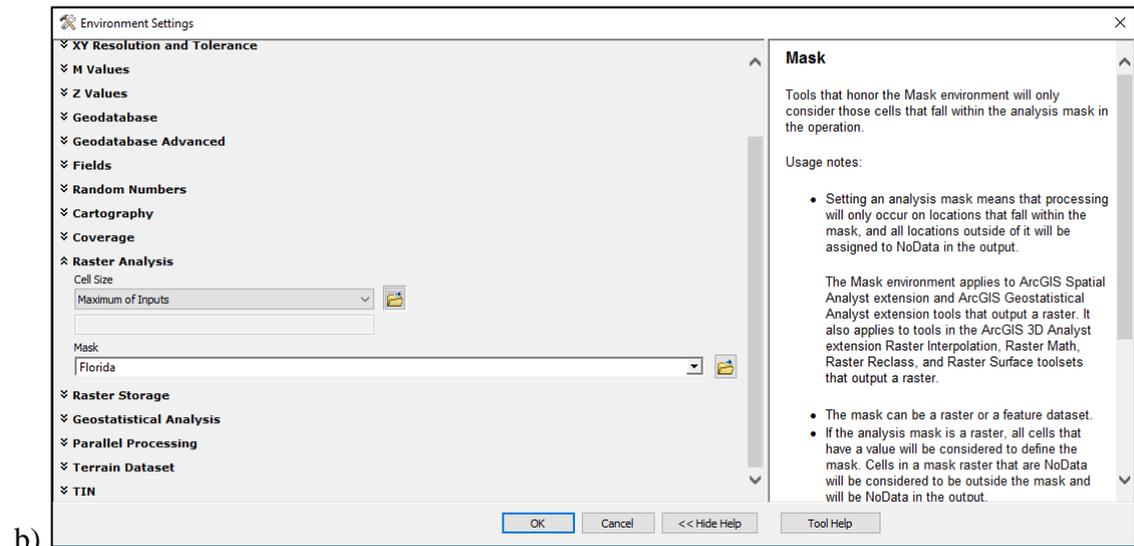
Figure 5. a) IDW tool with options filled in and b) the location of the mask setting.

An IDW raster will be created with the following characteristics: floating point pixel type, WGS 1984 geographic coordinate system, and a cell size of ~0.023. The default display is 9 equal intervals under classified symbology (Figure 6).
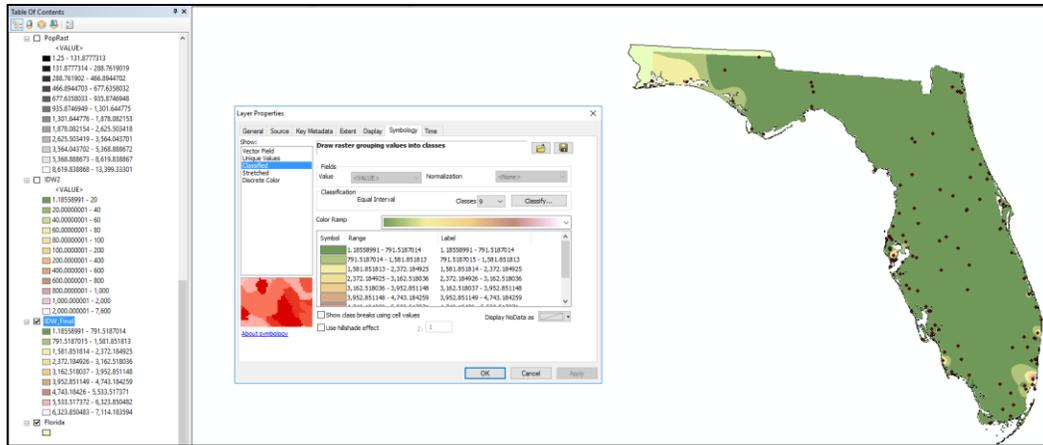
Figure 6. Initial IDW

If desired, change the intervals manually under classify to see a more representative spread of the data (Figure 7). I recommend the following values to account for the data distribution with a huge gap between ~2000 and 7000 starlings.
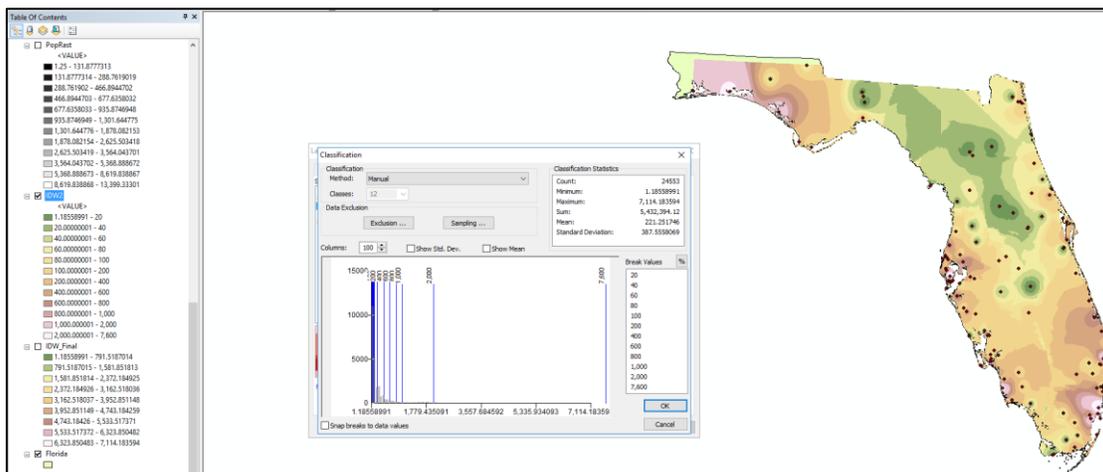
Figure 7. Starling IDW after the addition of manual classified intervals under symbology

Add the population data from ArcGIS Online by searching for USA Census Populated Places and choose the first option. This will give the populated locations throughout the United States (Figure 8). Clip the newly downloaded data to the Florida polygon with the Clip (Analysis) tool (Figure 9a) and change the symbology to display the population per square mile (POP12_SQMI; Figure 9b).
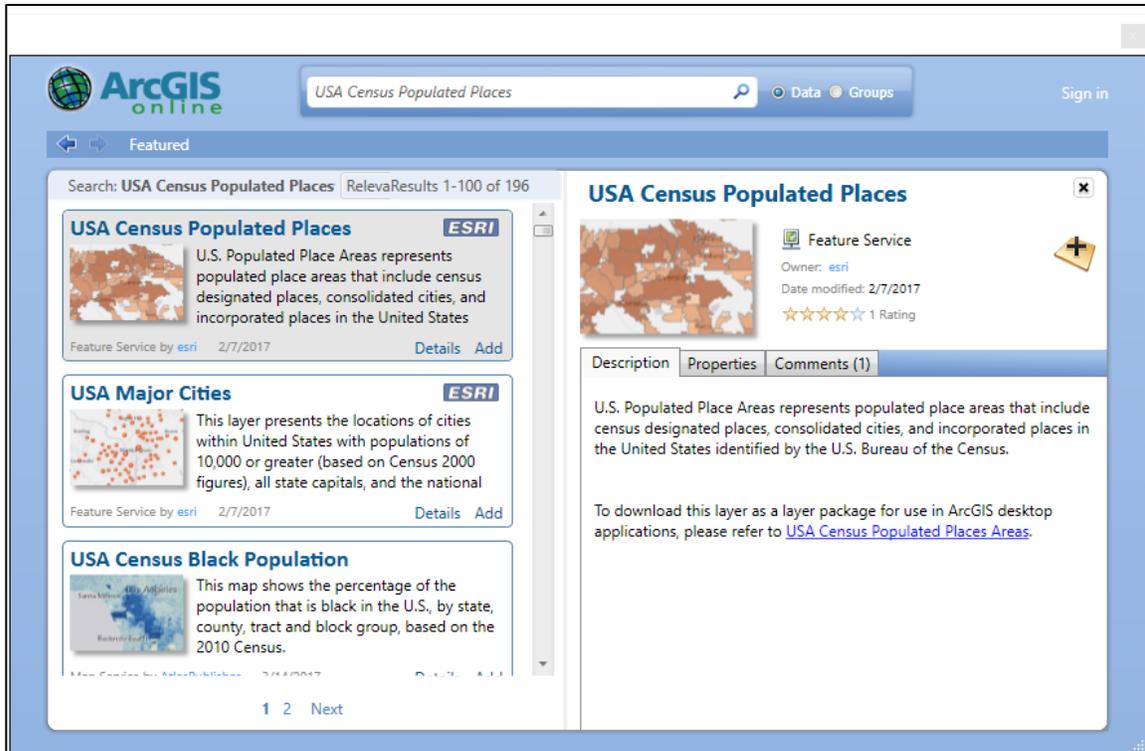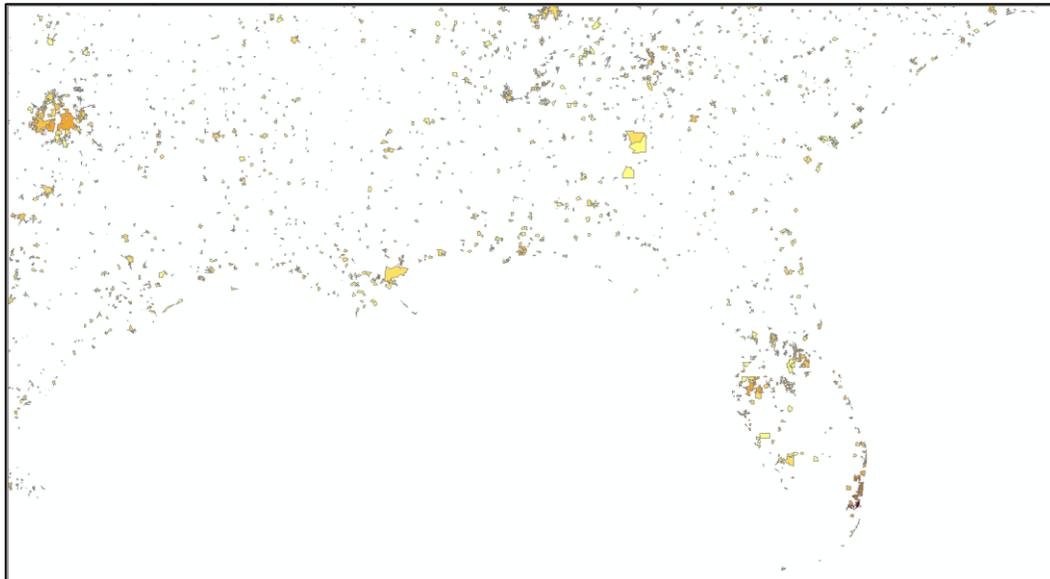
Figure 8. a) ArcGIS Online data search and b) a portion of the results
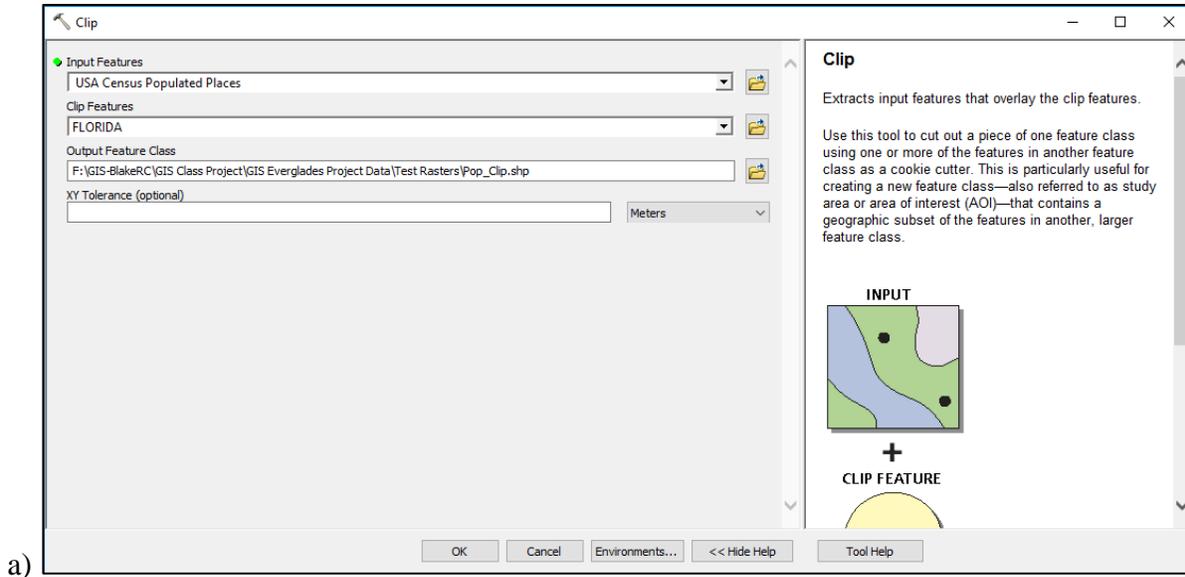
a)
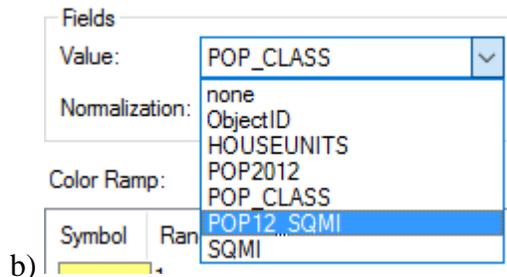
b)

Figure 9. The Clip tool (a) and the option to change the display symbology (b)

Convert the population per square mile clip to a raster using the Polygon to Raster (Conversion) tool (Figure 10). Choose population per square mile and a cell size of 2547 to create a similar cell size to the IDW raster (the cell size here was divided by 250). Convert the population raster and IDW raster to the NAD 1983 UTM Zone 17N with Project Raster (Data Management) tool, then upload to new ArcMap (Figure 11). Attempting to change the coordinate system on the raw data files will cause a misalignment of the data. The projected coordinate system of the Florida shapefile can be changed in ArcCatalog.
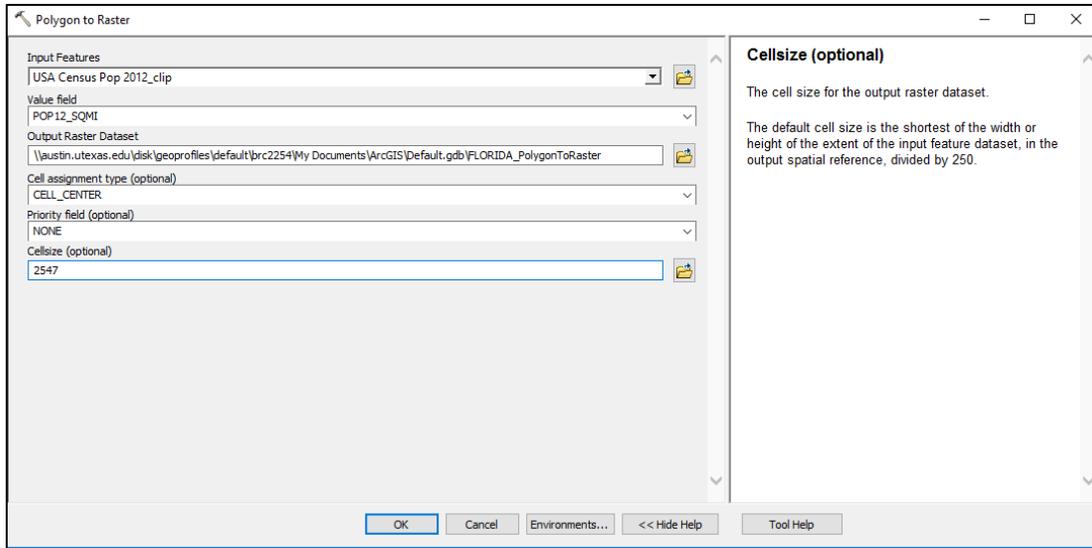
Figure 10. The Polygon to Raster tool
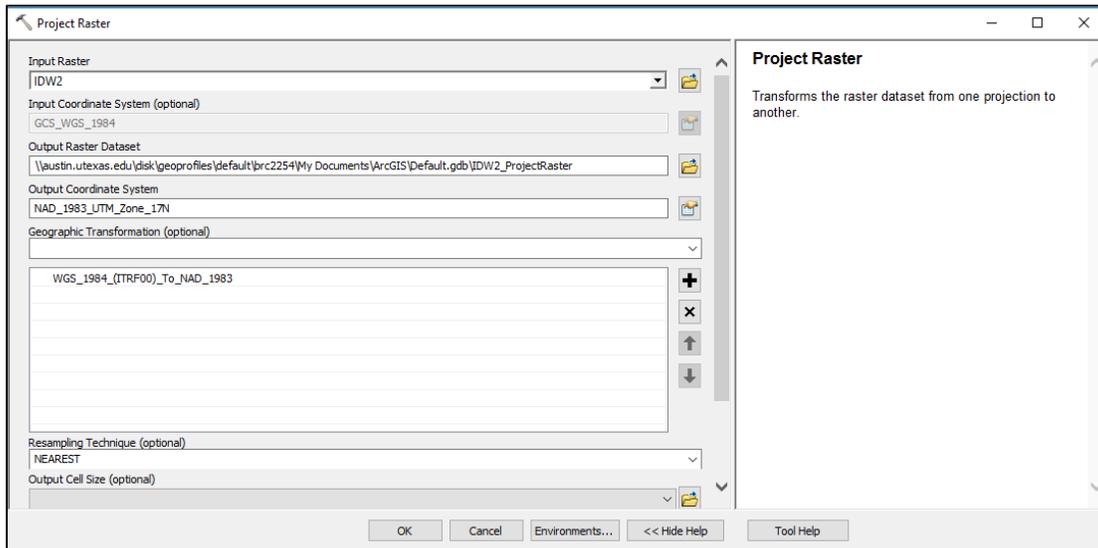
Figure 11. The Project Raster tool with the IDW raster within the input

The result resembles Figure 12 with a variable symbology. The values chosen for the IDW raster were intended to better represent the spread of data while the population raster divisions were chosen somewhat arbitrarily to create an even spread. Each was given 9 manually-chosen classified categories.

10
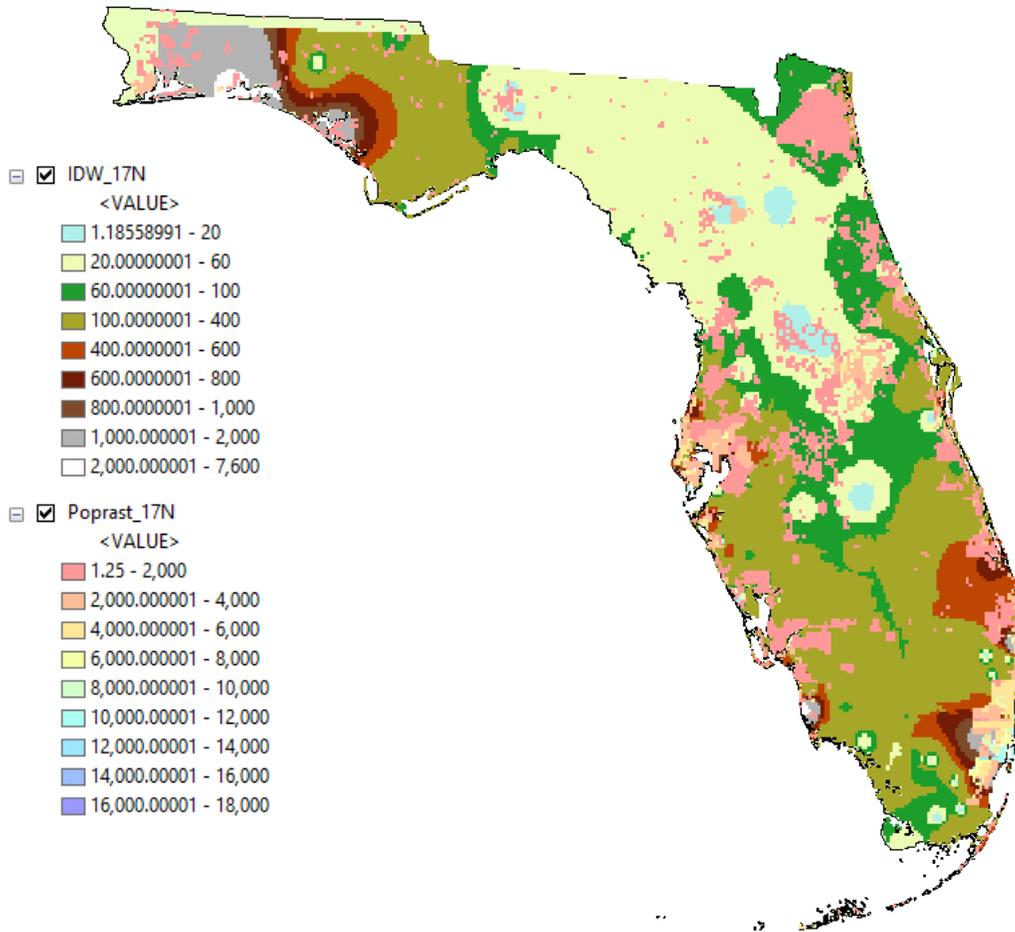
Figure 12. Results of the Project Raster tool when converting to NAD 1983 UTM Zone 17N

Reclassify both of your rasters with the Reclassify (Spatial Analyst) tool and leave the defaults (Figure 13a). You may find fewer categories in the output raster because of few data values at higher populations per square mile. The resulting rasters are shown below in Figure 13b, both with unsigned values.

a)

b)
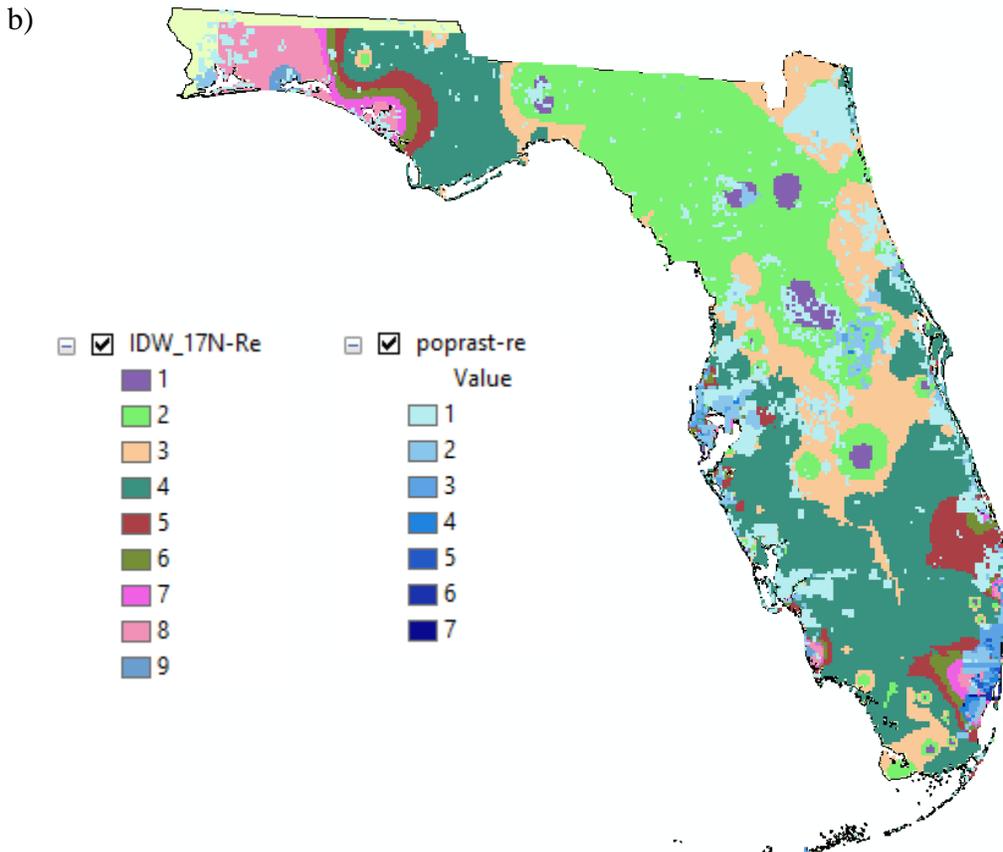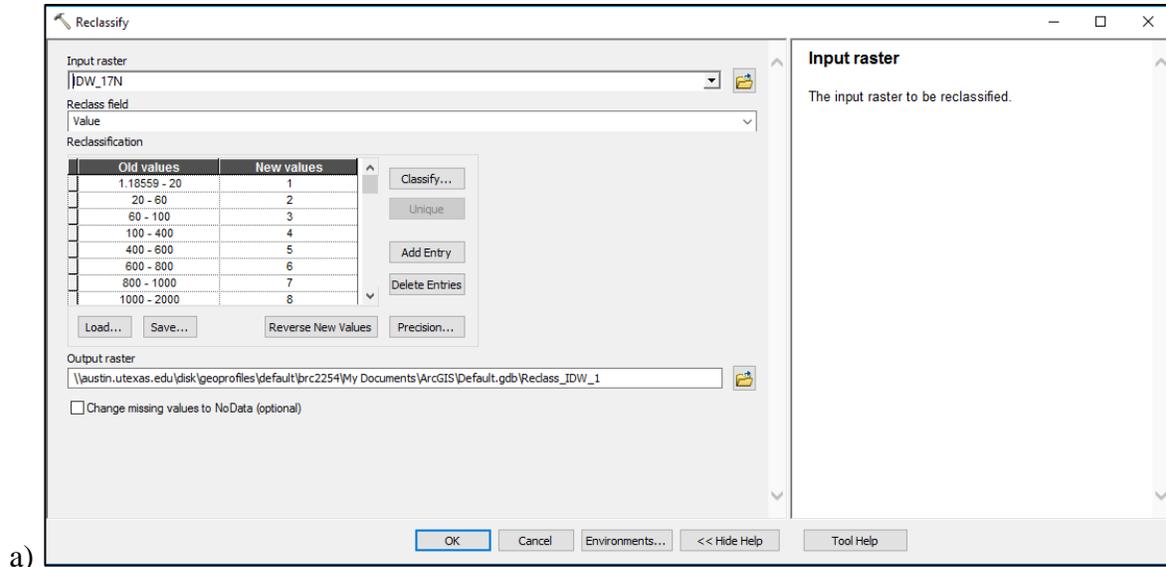
Figure 13. The raster Reclassify tool appears in a) and the results of the reclassification for the IDW and population rasters are shown in b)

The final step involves using the Raster Calculator to run the Pearson Correlation Coefficient equation (Figure 14). X corresponds to the first raster (I chose the reclassified

IDW raster) and its mean (found listed under the statistics in the properties menu) while Y shows the second raster (reclassified population raster). The correlation coefficient (r) ranges from -1 (strong negative correlation) to +1 (strong positive correlation) and can equal 0 (no correlation). After using the Raster Calculator, your result will appear like the final map (floating point pixel type). I recommend using the Cold to Hot Diverging symbology.

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2}\sqrt{\sum (Y - \overline{Y})^2}}$$

("Computing")
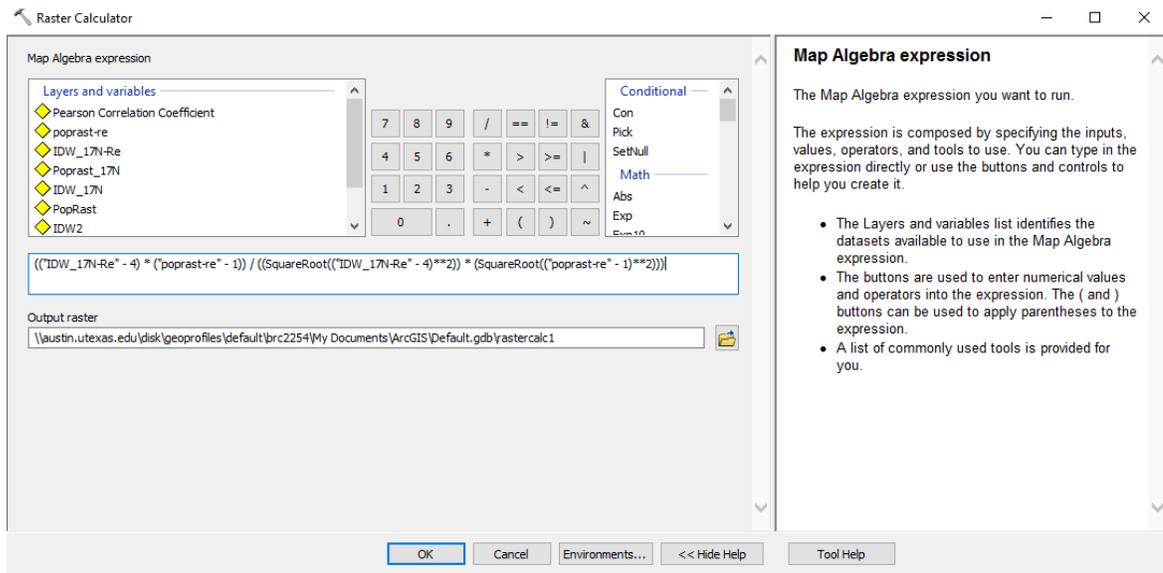
Figure 14. The Pearson Correlation Coefficient equation formatted within the Raster Calculator

IV. Discussion

The final map only appeared with +1 (red) and -1 (blue) values seemingly indicating regions of strong positive (high starling count and high population densities) and strong negative (likely low starling count and high population densities) correlations. In the context of the data used to fill the Pearson correlation coefficient, the equation will only tell me that the starlings and population densities of humans will be correlated above (+1) or below (-1) their respective mean values. One caveat is that slight differences from the mean will cause the end-member values with no middle values (correlation regions are surrounded by areas labeled as no data). If one variable is a little under the mean and the other slightly above, then we would expect -1. If both are a little above or below the mean, then we expect +1. The denominator is always positive, acting as an absolute value. The Pearson correlation coefficient is meant for changes through time that are summed together and is perhaps not the most appropriate correlation technique for this analysis. Further research could uncover a new method that may lead to more reliable results.

13

V. References

Cabe, Paul R. 1993. European Starling (*Sturnus vulgaris*), version 2.0. In The Birds of North America (P. G. Rodewald, editor). Cornell Lab of Ornithology, Ithaca, New York, USA. https://doi.org/10.2173/bna.48.

"Computing the Pearson Correlation Coefficient," Western Michigan University, 2003: http://www.stat.wmich.edu/s216/book/node122.html (accessed December 2017).

EDDMapS, 2017, University of Georgia, Distribution Maps: https://www.eddmaps.org/distribution/ (accessed November 2017).

Linz, G.M., Homan, H.J., Gaukler, S.M., Penry, L.B., and Bleier, W.J., 2007, European starlings: a review of an invasive species with far-reaching impacts: Managing Vertebrate Invasive Species: Proceedings of an International Symposium, p. 378-386.

"Nonnatives - Euopean Starling [sic]," Florida Fish and Wildlife Conservation Commission: http://myfwc.com/wildlifehabitats/nonnatives/birds/european-starling/ (accessed December 2017).

Pearson, K., 1920, Notes on the History of Correlation: Biometrika, v. 13, no. 1, p. 25-45.

Southwest Florida Water Management District's Geospatial Open Data Portal, 2017: http://data-swfwmd.opendata.arcgis.com/ (accessed November 2017).

USA Census 2012 data collected by the United States Census Bureau and compiled for ArcGIS Online.

Blake Chapman
December 07, 2017

Correlation of Invasive European Starlings and High Human Population per Square Mile in 2012 in Florida by the Pearson Coefficient

Explanation

Pearson Correlation Coefficient

High : 1

Low : -1

Florida

0   25   50
Kilometers

N

NAD 1983 UTM Zone 17N