GIS: Geographic Information Systems

Module 11: Basic spatial modelling

Matthew L. Sisk
Center for Digital Scholarship
Hesburgh Library, University of Notre Dame

library.nd.edu/cds/

Spatial analysis:

A set of methods whose results change when the locations of the objects being analyzed changes

Longley et al. 2005:

Spatial analysis is deriving information from data using distributions or patterns

Spatial modeling

- A model is a representation of reality
  - Created as a simplified, manageable view of reality
  - Help understand, describe or predict how things work in the real world

Spatial analysis examples

CENTER for DIGITAL SCHOLARSHIP

- Vector
  - Interactive selections
  - Spatial joins
- Zonal statistics
- Raster
  - Distance and space allocation
  - Cost weighted distance
  - Raster Algebra and complex calculations

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Spatial joins

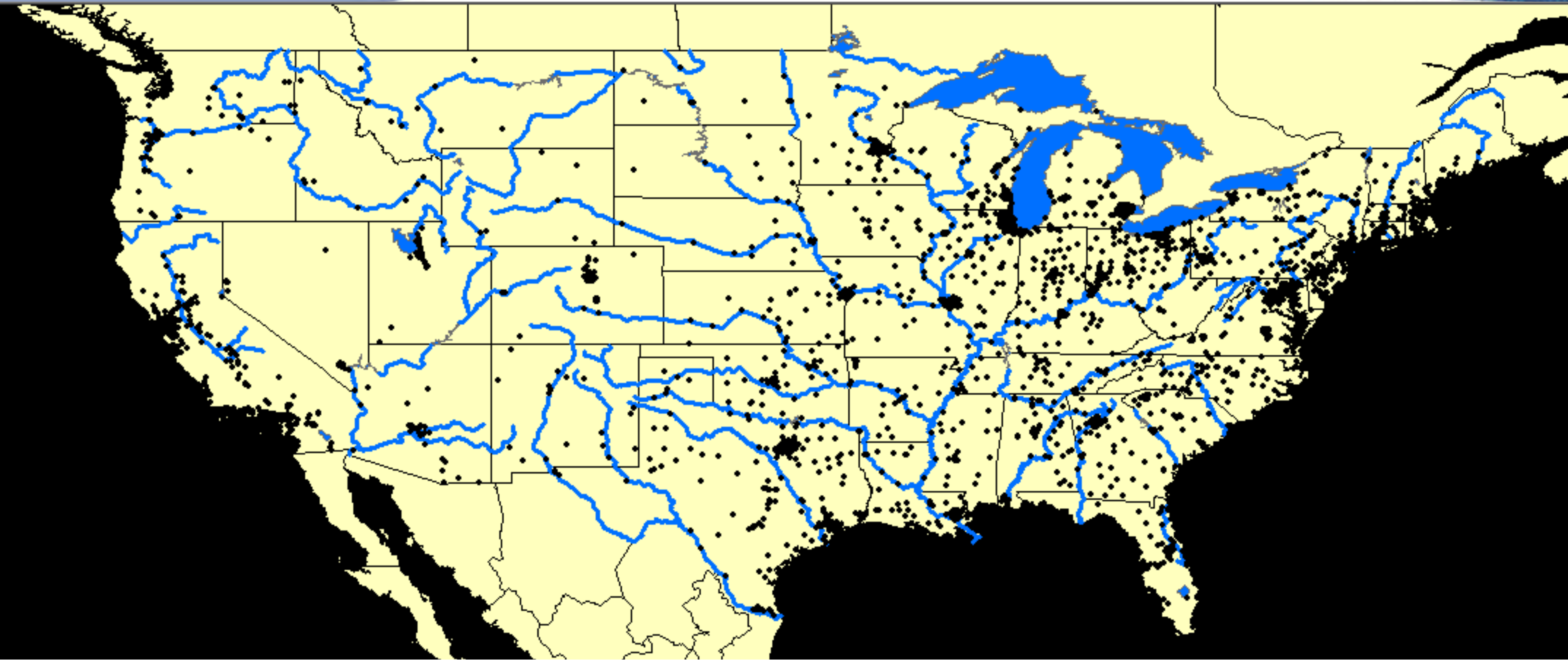
CENTER for DIGITAL SCHOLARSHIP

- Adds information from another shapefile based on spatial location
- Can calculate based on most of the selections (distance, intersections, etc) outlined in Module 3

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Spatial Join Example

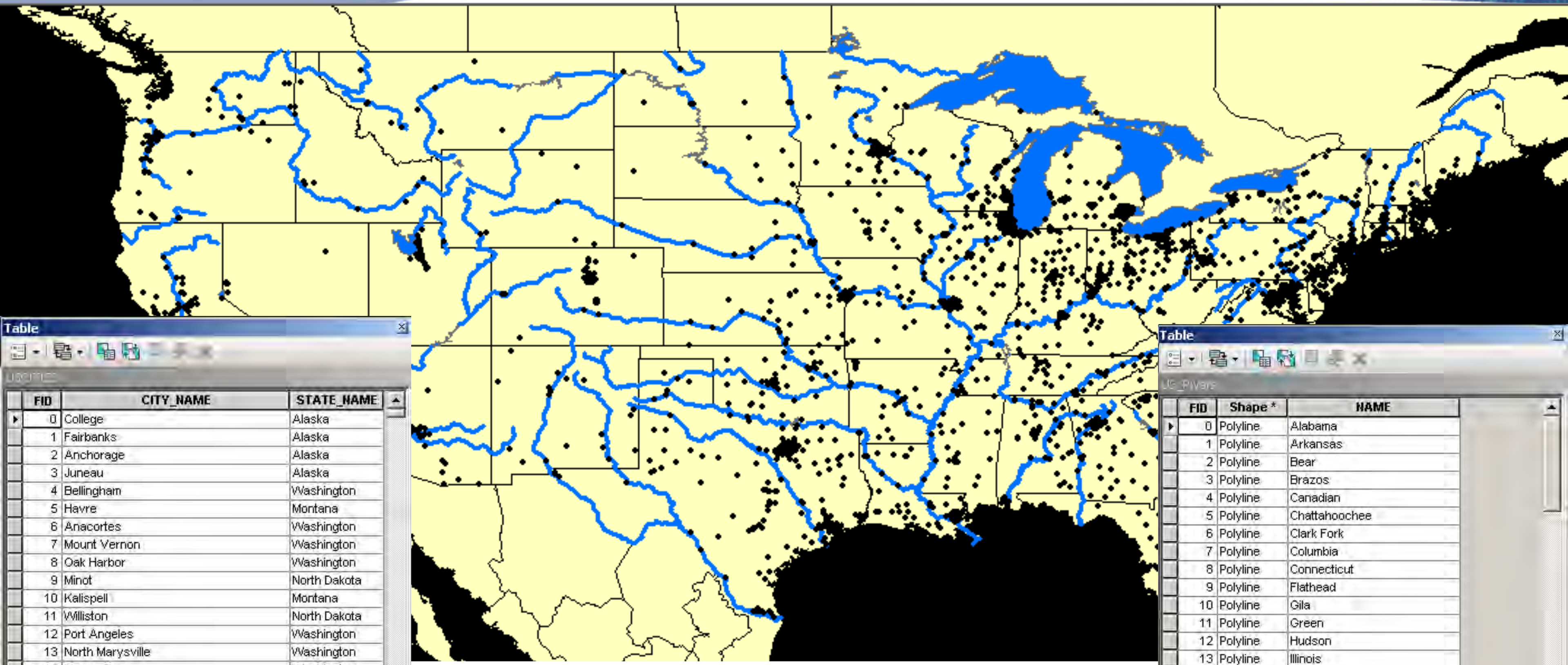
CENTER for DIGITAL SCHOLARSHIP



Add fields to the cities shapefile showing the nearest river and its distance.

UNIVERSITY OF NOTRE DAME Hesburgh Libraries

Spatial Join Example



Table

USCITIES

FID	CITY_NAME	STATE_NAME
0	College	Alaska
1	Fairbanks	Alaska
2	Anchorage	Alaska
3	Juneau	Alaska
4	Bellingham	Washington
5	Havre	Montana
6	Anacortes	Washington
7	Mount Vernon	Washington
8	Oak Harbor	Washington
9	Minot	North Dakota
10	Kalispell	Montana
11	Williston	North Dakota
12	Port Angeles	Washington
13	North Marysville	Washington
14	Marysville	Washington
15	West Lake Stevens	Washington
16	Everett	Washington
17	Grand Forks	North Dakota
18	Paine Field-Lake Stickney	Washington

1 (0 out of 3149 Selected)

USCITIES

US_Cities Shapefile

Table

US_Rivers

FID	Shape *	NAME
0	Polyline	Alabama
1	Polyline	Arkansas
2	Polyline	Bear
3	Polyline	Brazos
4	Polyline	Canadian
5	Polyline	Chattahoochee
6	Polyline	Clark Fork
7	Polyline	Columbia
8	Polyline	Connecticut
9	Polyline	Flathead
10	Polyline	Gila
11	Polyline	Green
12	Polyline	Hudson
13	Polyline	Illinois
14	Polyline	Kings
15	Polyline	Klamath
16	Polyline	Mississippi
17	Polyline	North Platte
18	Polyline	Ohio

1 (0 out of 61 Selected)

USCITIES US_Rivers

US_Rivers Shapefile

Spatial Join Example

Table

USCITIES

FID	CITY_NAME	STATE_NAME
0	College	Alaska
1	Fairbanks	Alaska
2	Anchorage	Alaska
3	Juneau	Alaska
4	Bellingham	Washington
5	Havre	Montana
6	Anacortes	Washington
7	Mount Vernon	Washington
8	Oak Harbor	Washington
9	Minot	North Dakota
10	Kalispell	Montana
11	Williston	North Dakota
12	Port Angeles	Washington
13	North Marysville	Washington
14	Marysville	Washington
15	West Lake Stevens	Washington
16	Everett	Washington
17	Grand Forks	North Dakota
18	Paine Field-Lake Stickney	Washington

1 (0 out of 3149 Selected)

USCITIES

US_Cities Shapefile

Table

US_Rivers

FID	Shape *	NAME
0	Polyline	Alabama
1	Polyline	Arkansas
2	Polyline	Bear
3	Polyline	Brazos
4	Polyline	Canadian
5	Polyline	Chattahoochee
6	Polyline	Clark Fork
7	Polyline	Columbia
8	Polyline	Connecticut
9	Polyline	Flathead
10	Polyline	Gila
11	Polyline	Green
12	Polyline	Hudson
13	Polyline	Illinois
14	Polyline	Kings
15	Polyline	Klamath
16	Polyline	Mississippi
17	Polyline	North Platte
18	Polyline	Ohio

1 (0 out of 61 Selected)

USCITIES US_Rivers

US_Rivers Shapefile

Join Data

Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.

What do you want to join to this layer?

Join data from another layer based on spatial location

1. Choose the layer to join to this layer, or load spatial data from disk:

US_Rivers

2. You are joining: Lines to Points

Select a join feature class above. You will be given different options based on geometry types of the source feature class and the join feature class.

- Each point will be given a summary of the numeric attributes of the lines that intersect it, and a count field showing how many lines intersect it.

  How do you want the attributes to be summarized?

  - Average
  - Minimum
  - Standard Deviation
  - Sum
  - Maximum
  - Variance

- Each point will be given all the attributes of the line that is closest to it, and a distance field showing how close that line is (in the units of the target layer).

3. The result of the join will be saved into a new layer.

Specify output shapefile or feature class for this new layer:

Y:\OpenProjects\Dissertation\GISData\Classifications\Corine

About Joining Data | OK | Cancel

Table

Citys_NearestWater

FID	US_Rivers_FID	CITY_NAME	STATE_NAME	FID_2	NAME	Distance
267	2160	Bethany	Oklahoma	4	Canadian	0.183884
268	2165	Oklahoma City	Oklahoma	4	Canadian	0.162207
269	2166	Midwest City	Oklahoma	4	Canadian	0.236949
270	2168	Del City	Oklahoma	4	Canadian	0.176871
271	2172	Elk City	Oklahoma	4	Canadian	0.496503
272	2173	Mustang	Oklahoma	4	Canadian	0.061485
273	2177	Shawnee	Oklahoma	4	Canadian	0.389049
274	2181	Moore	Oklahoma	4	Canadian	0.077813
275	2193	Norman	Oklahoma	4	Canadian	0.096667
276	2219	Chickasha	Oklahoma	4	Canadian	0.316459
277	2232	McAlester	Oklahoma	4	Canadian	0.266396
278	2249	Ada	Oklahoma	4	Canadian	0.107997
279	2288	Gainesville	Georgia	5	Chattahoo	0.036514
280	2355	Alpharetta	Georgia	5	Chattahoo	0.090317
281	2372	Roswell	Georgia	5	Chattahoo	0.045049
282	2408	Athens	Georgia	5	Chattahoo	0.556373
283	2409	Marietta	Georgia	5	Chattahoo	0.126174
284	2410	Lawrenceville	Georgia	5	Chattahoo	0.160103
285	2413	Dunwoody	Georgia	5	Chattahoo	0.038596
286	2418	Sandy Springs	Georgia	5	Chattahoo	0.033802

1 (0 out of 3149 Selected)

Citys_NearestWater

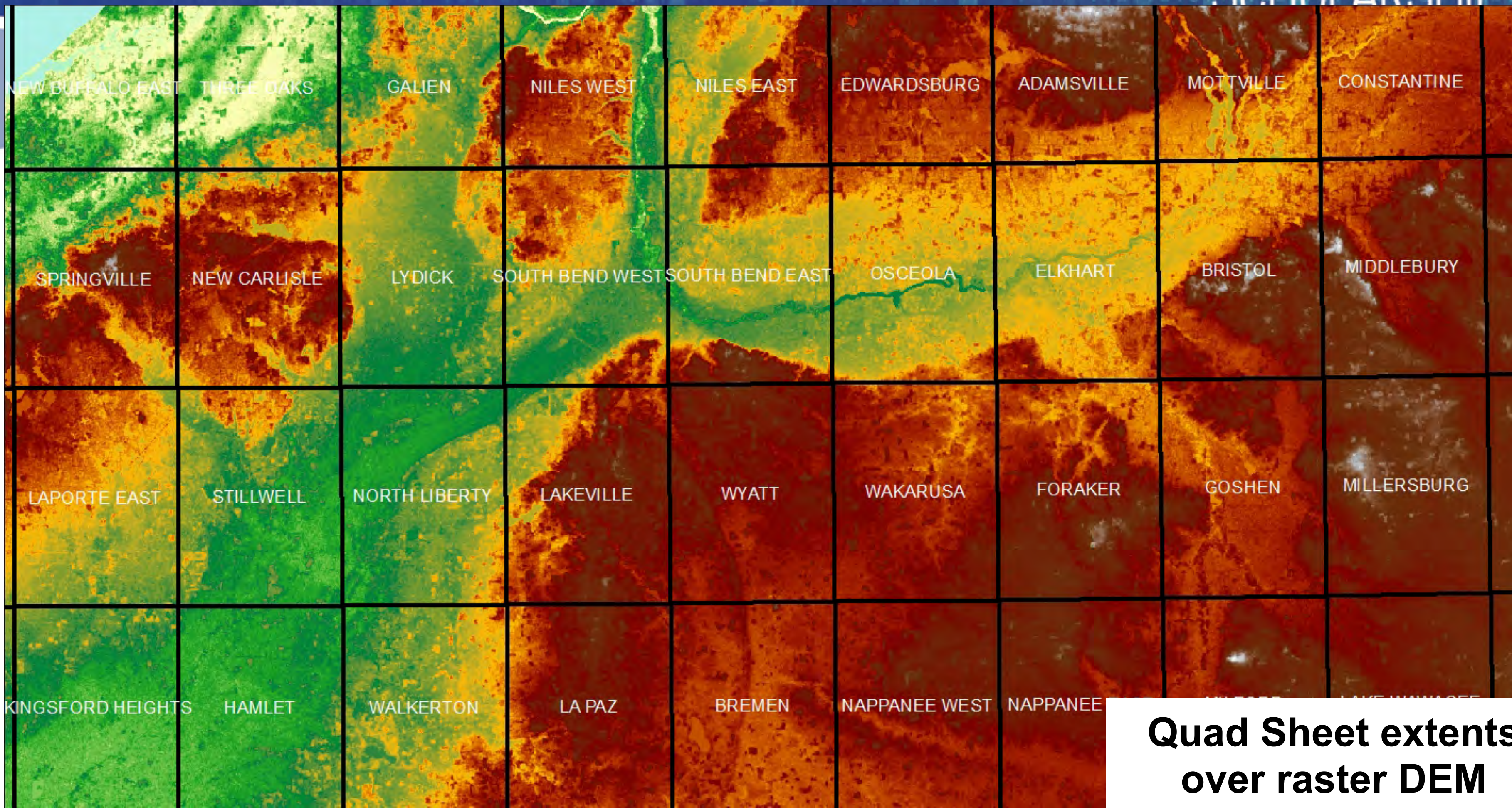
New Shapefile with nearest river and distance fields

Zonal Statistics

- Summarizes the variation in a raster for given areas
  - These areas can from a polygon shapefile or classified raster image
- Output can be a raster image or a summary table

Zonal Statistics Example

CENTER for DIGITAL SCHOLARSHIP



Quad Sheet extents over raster DEM

Zonal Statistics Example

CENTER for DIGITAL SCHOLARSHIP

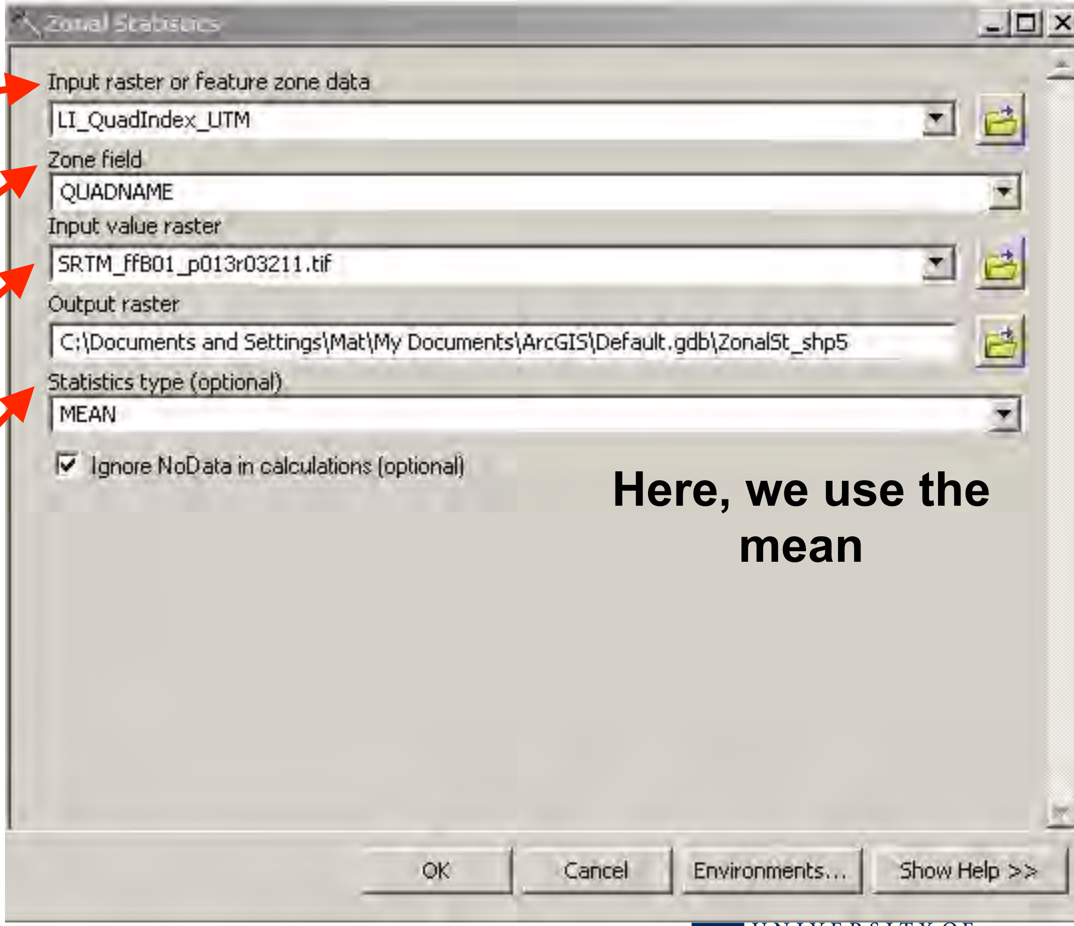
Specify:

Zonal data

Field to use in reports

Raster to summarize

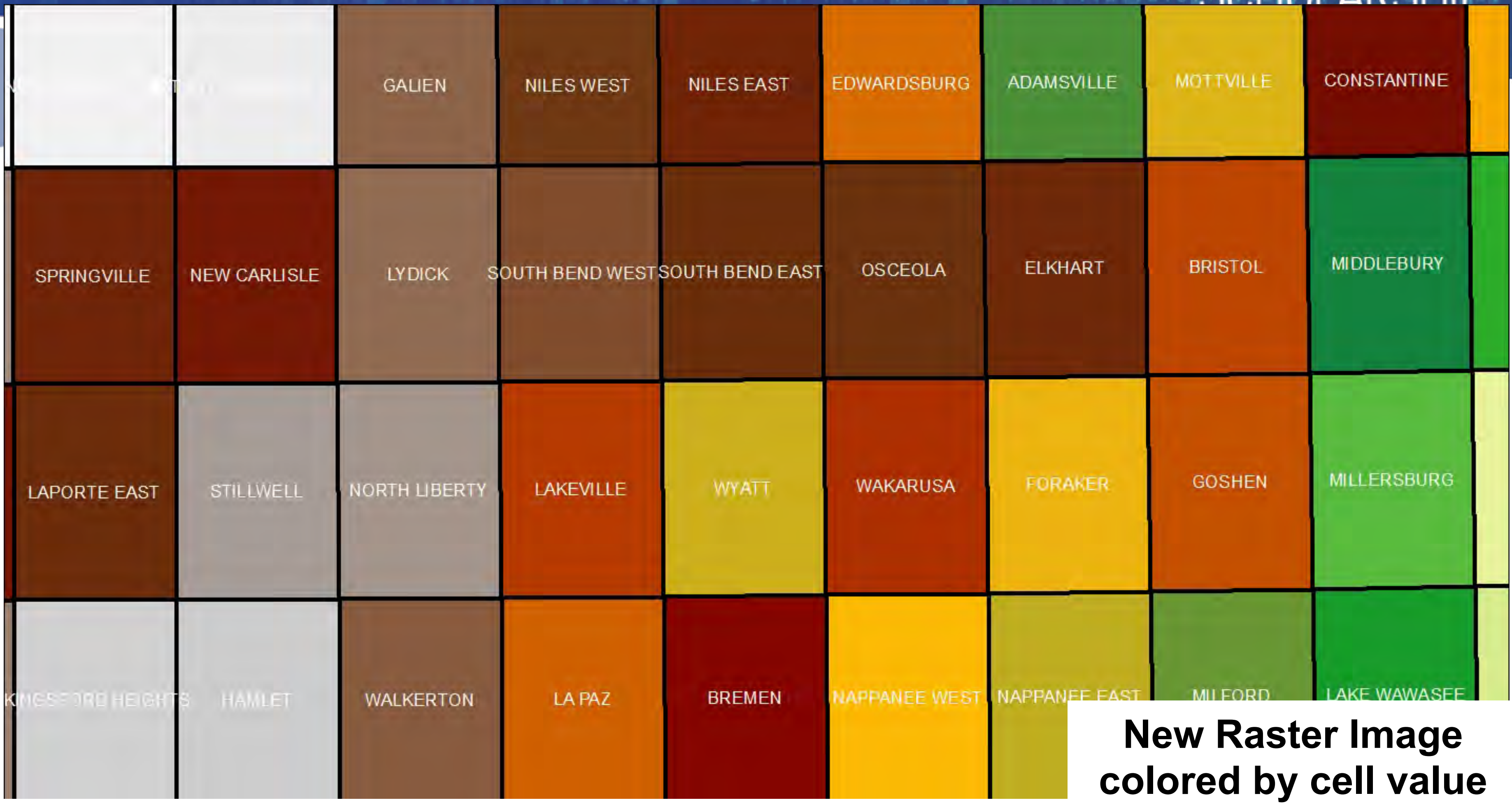
The measure to report



Here, we use the mean

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Zonal Statistics Example



New Raster Image colored by cell value

Zonal Statistics Example

CENTER for DIGITAL SCHOLARSHIP

Table

ZonalSt_shp3

OBJECTID *	NAME	ZONE_CODE	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM	VARIETY	MAJORITY	MINORITY	MEDIAN
16	ANGOLA WEST	16	197064	144450900	283	351	68	306.4791	11.75023	6039600	69	290	351	306
17	ORLAND	17	197065	144451600	272	338	66	294.2527	9.033888	5798691	67	289	337	292
18	MONGO	18	197046	144437700	261	324	63	284.5303	9.164295	5606556	64	278	261	284
19	LAGRANGE	19	197031	144426700	253	312	59	271.5516	9.071217	5350409	59	269	312	270
20	SHIPSHEWANA	20	197000	144404000	247	306	59	267.6978	8.542318	5273647	60	258	306	268
21	MIDDLEBURY	21	197005	144407600	233	314	81	263.6422	11.36197	5193883	81	259	233	263
22	BRISTOL	22	197008	144409800	225	304	79	251.729	15.10875	4959263	80	235	225	249
23	ELKHART	23	196981	144390000	216	269	53	236.2276	6.250952	4653235	54	234	216	235
24	OSCEOLA	24	196979	144388600	214	261	47	232.663	6.863205	4582972	48	231	261	232
25	SOUTH BEND EAST	25	196991	144397300	197	290	93	234.351	14.33497	4616504	93	228	197	230
26	SOUTH BEND WEST	26	196952	144368800	190	276	86	228.9691	12.26284	4509592	87	218	192	227
27	LYDICK	27	196962	144376100	209	287	78	223.9474	8.498893	4410914	79	221	209	222
28	NEW CARLISLE	28	196934	144355600	203	275	72	245.0783	9.910855	4826424	73	246	275	246
29	SPRINGVILLE	29	196993	144398800	192	302	110	237.9041	21.82382	4686544	111	248	192	243
30	MICHIGAN CITY EAST	30	196991	144397300	175	300	125	219.5174	27.13097	4324295	126	202	299	208
31	MICHIGAN CITY WEST	31	196821	144272700	151	238	87	191.5984	13.16745	3771059	80	175	151	195
32	ASHLEY	32	197467	144746300	281	347	66	302.4638	9.463956	5972662	67	299	281	301
33	STROH	33	197431	144719900	271	338	67	300.2797	11.22851	5928451	68	308	273	301
34	WOLCOTTVILLE	34	197430	144719200	267	334	67	296.2541	7.776434	5848944	68	295	332	296
35	OLIVER LAKE	35	197398	144695700	263	312	49	281.5159	7.425364	5557068	50	285	263	282

1 (0 out of 111 Selected)

ZonalSt_shp3

Can also generate a summary table. This can then be joined back to the shapefile.

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

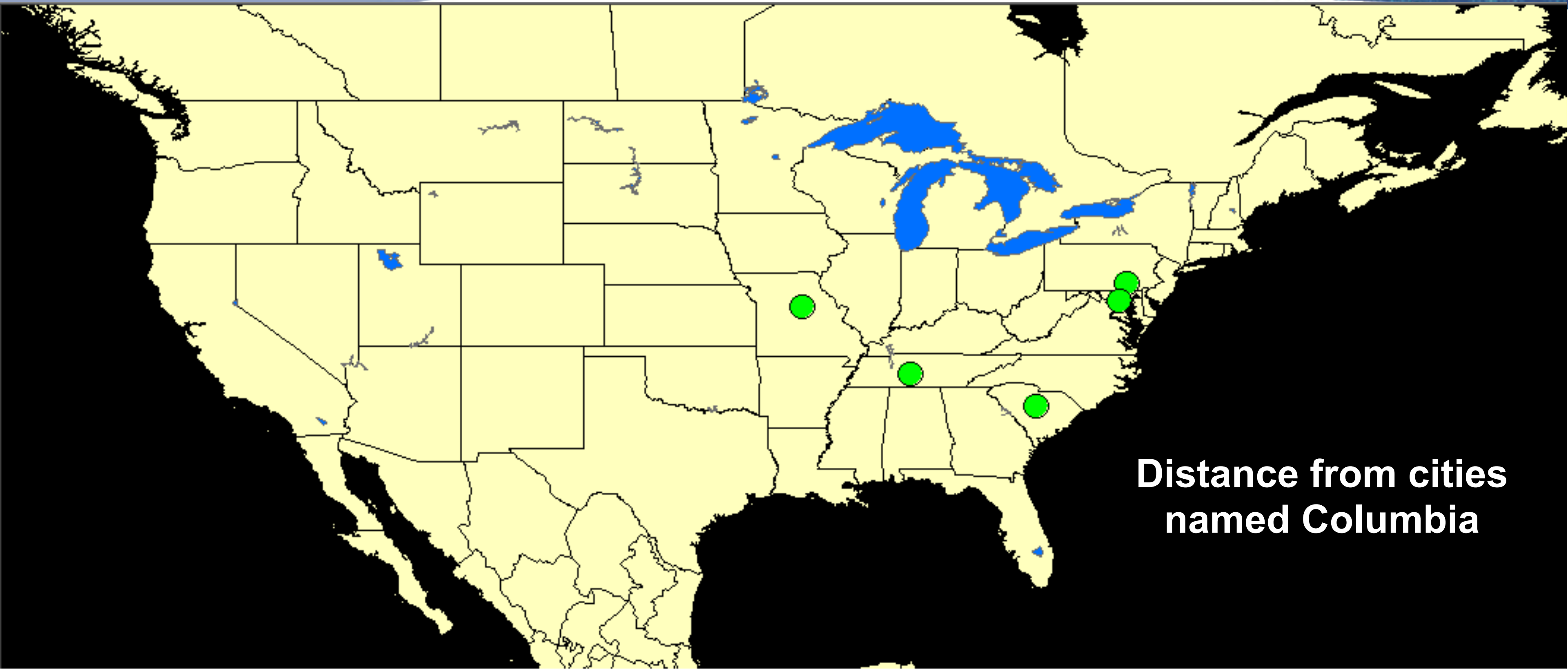
Euclidean Distance

CENTER for DIGITAL SCHOLARSHIP

- The straight-line distance between two points on a plane calculated using the Pythagorean theorem
- "As the crow flies"
- Creates a raster image where the only value for each cell is the distance from the feature being measured

Euclidean Distance

CENTER for DIGITAL SCHOLARSHIP

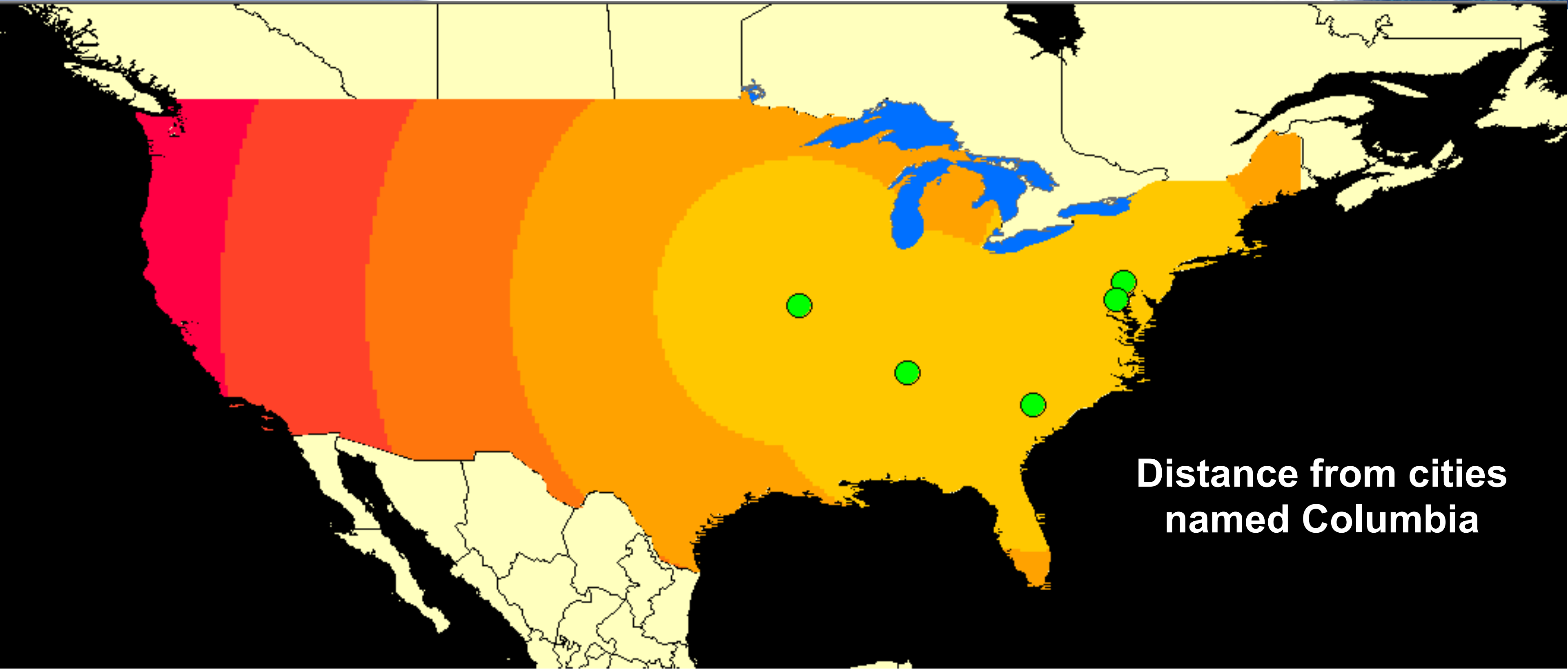


Yields a raster with distance from the given feature as the only pixel value

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Euclidean Distance

CENTER for DIGITAL SCHOLARSHIP



Yields a raster with distance from the given feature as the only pixel value

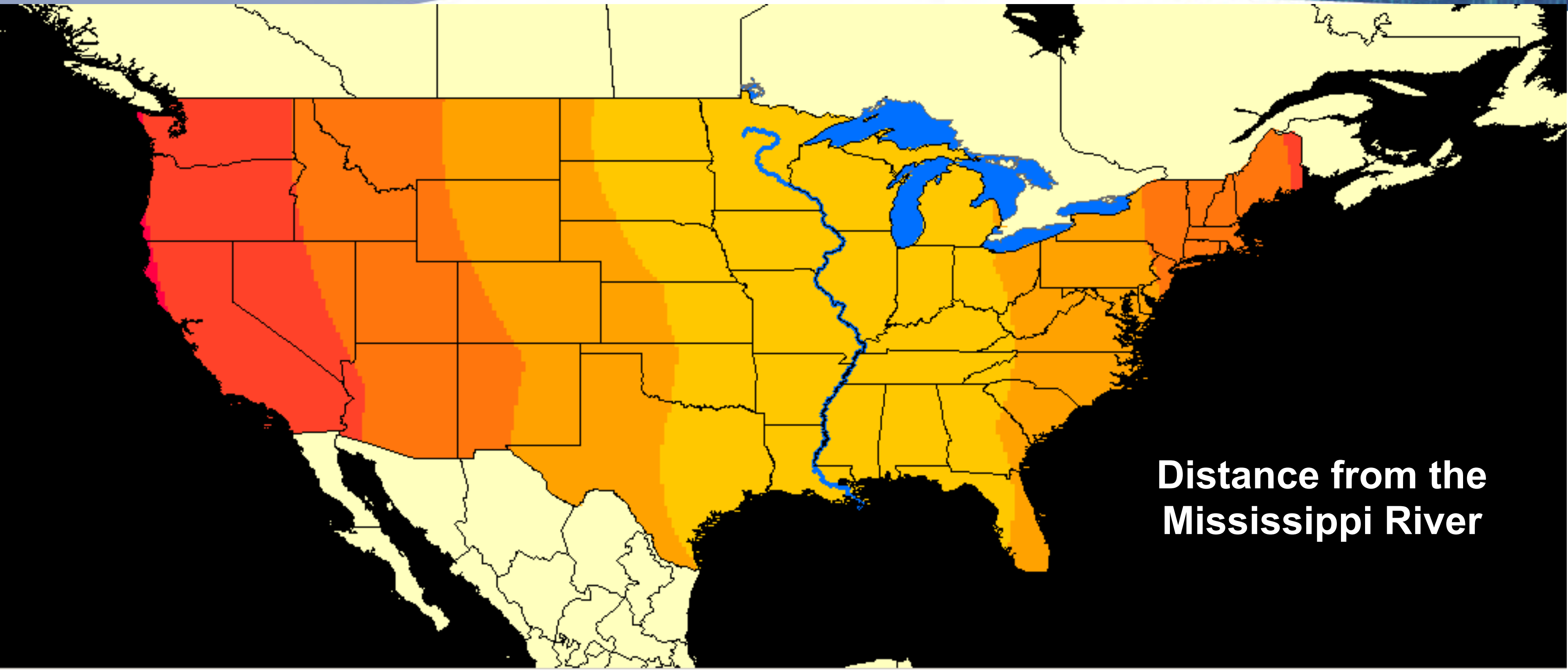
UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Euclidean Distance

Yields a raster with distance from the given feature as the only pixel value

Euclidean Distance

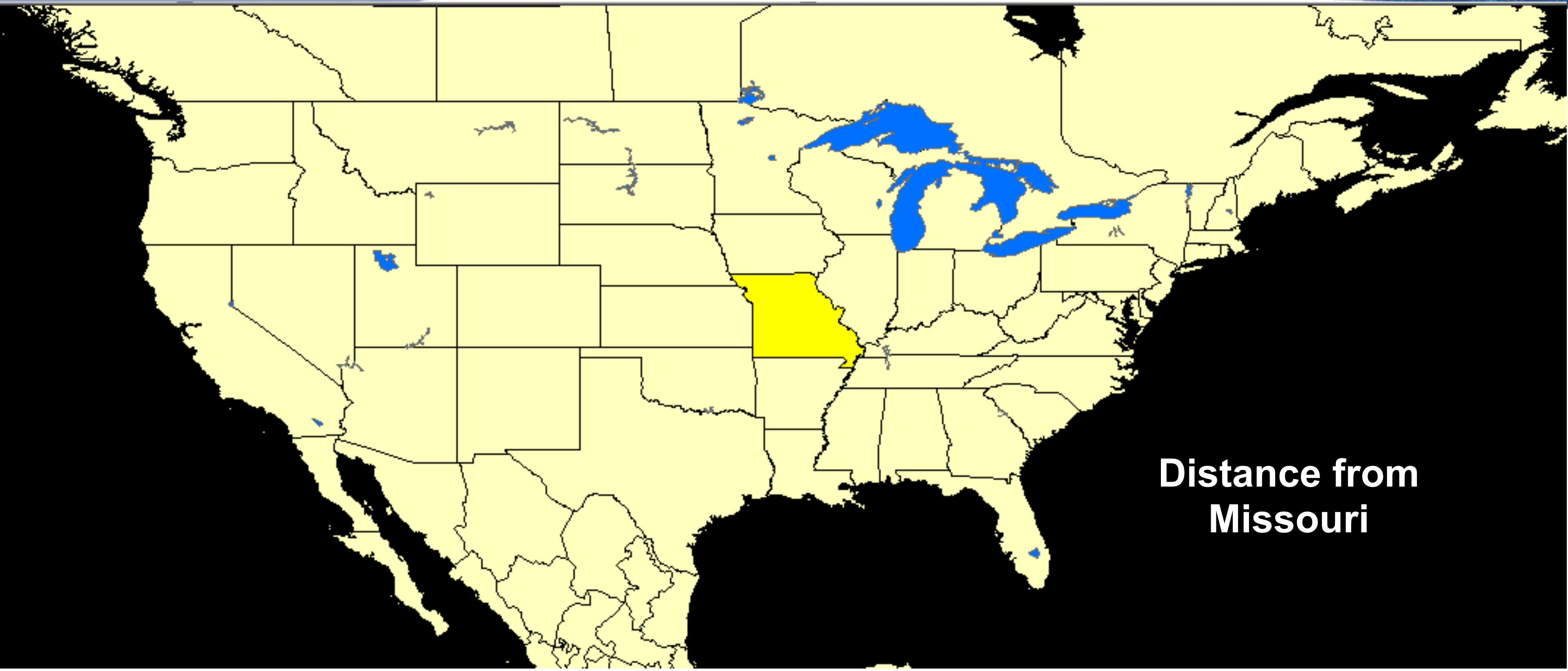
CENTER for DIGITAL SCHOLARSHIP



Yields a raster with distance from the given feature as the only pixel value

UNIVERSITY OF NOTRE DAME Hesburgh Libraries

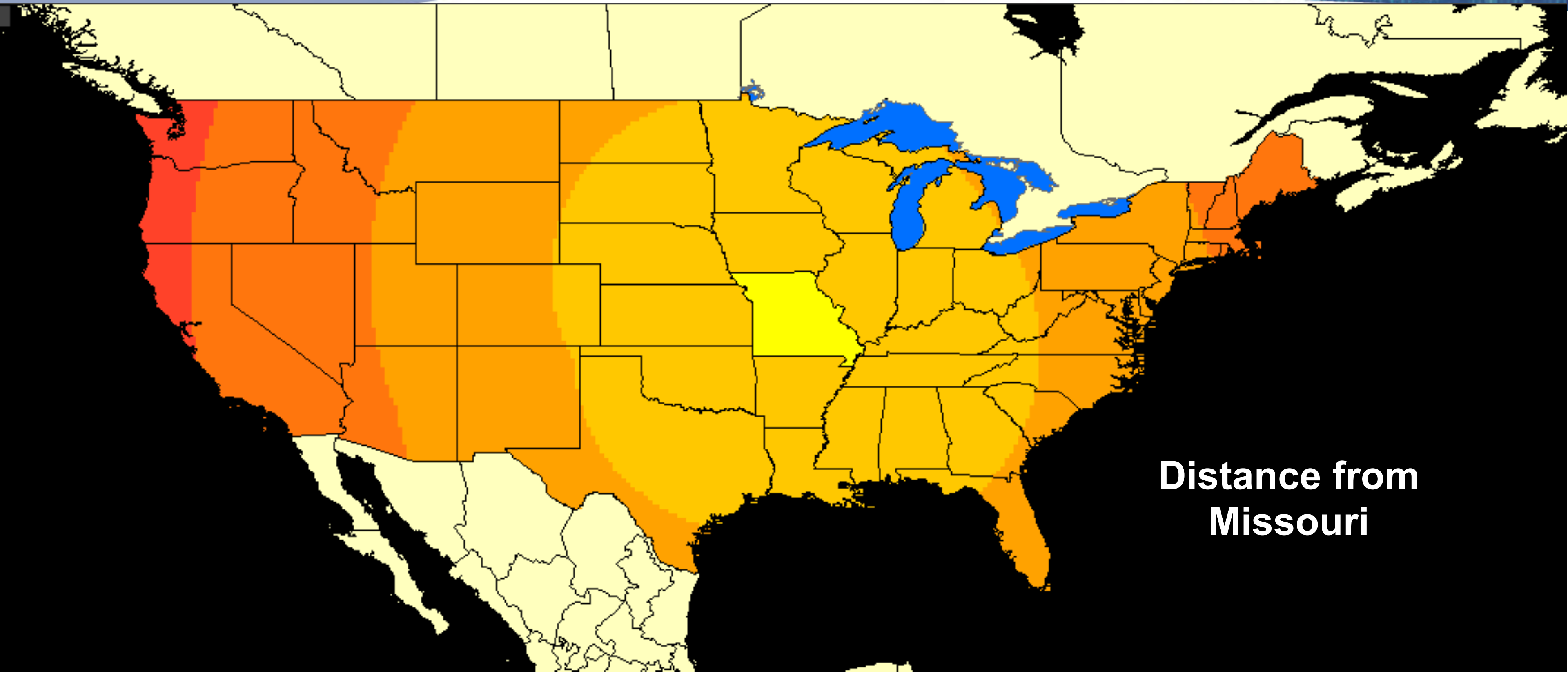
Euclidean Distance



Yields a raster with distance from the given feature as the only pixel value

Euclidean Distance

CENTER for DIGITAL SCHOLARSHIP



Yields a raster with distance from the given feature as the only pixel value

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

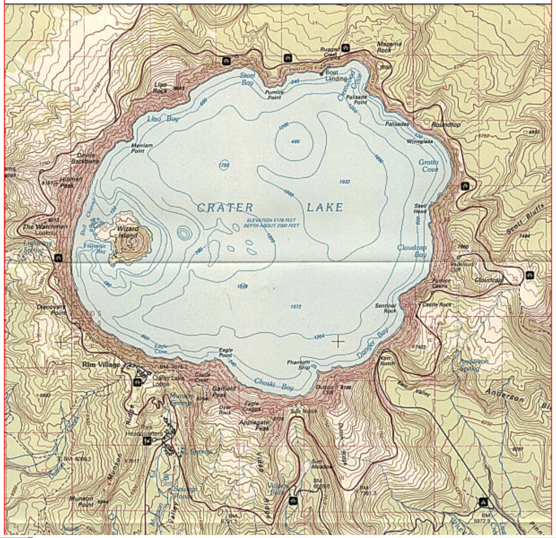
Cost Weighted Distance

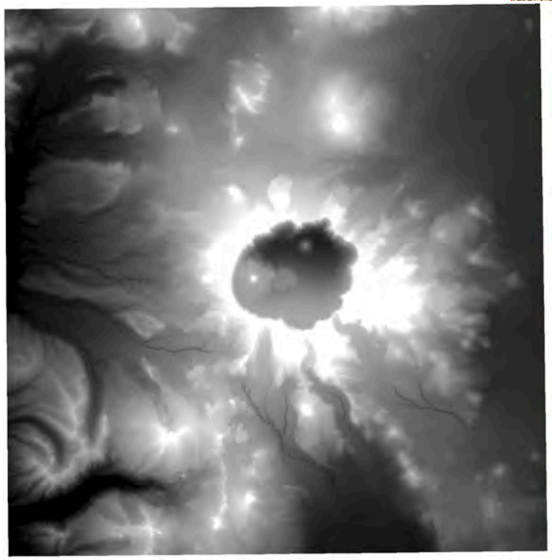
- In reality, the Euclidean distance is often not an important component of how easy it is to travel
  - Thus, we can use another layer to “weight” how easy it is to travel
- **Cost Raster**: A raster dataset that indicates the cost of traveling through each cell in the raster
  - Can be used to create a cumulative cost of traveling from each cell in the raster to a certain destination

Cost Weighted Distance

- For example, a route based on the Euclidean distance from one side of Crater Lake to the other would traverse the lake
- In reality, it would be much easier for someone to travel around the rim
- A longer distance, but a much easier route

Crater Lake, Oregon, Topographic Map



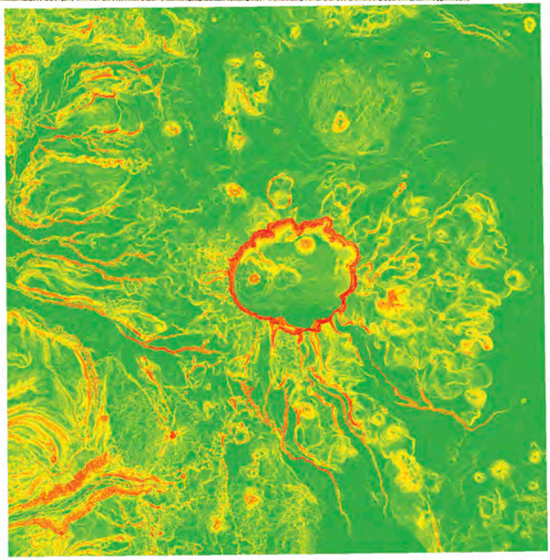


Crater Lake DEM

Elevation (m)

2323.7

945.4



Slope Degrees

0 - 4

4.0 1- 8

8.01 - 12

12.01 - 18

18.01 - 24

24.01 - 30

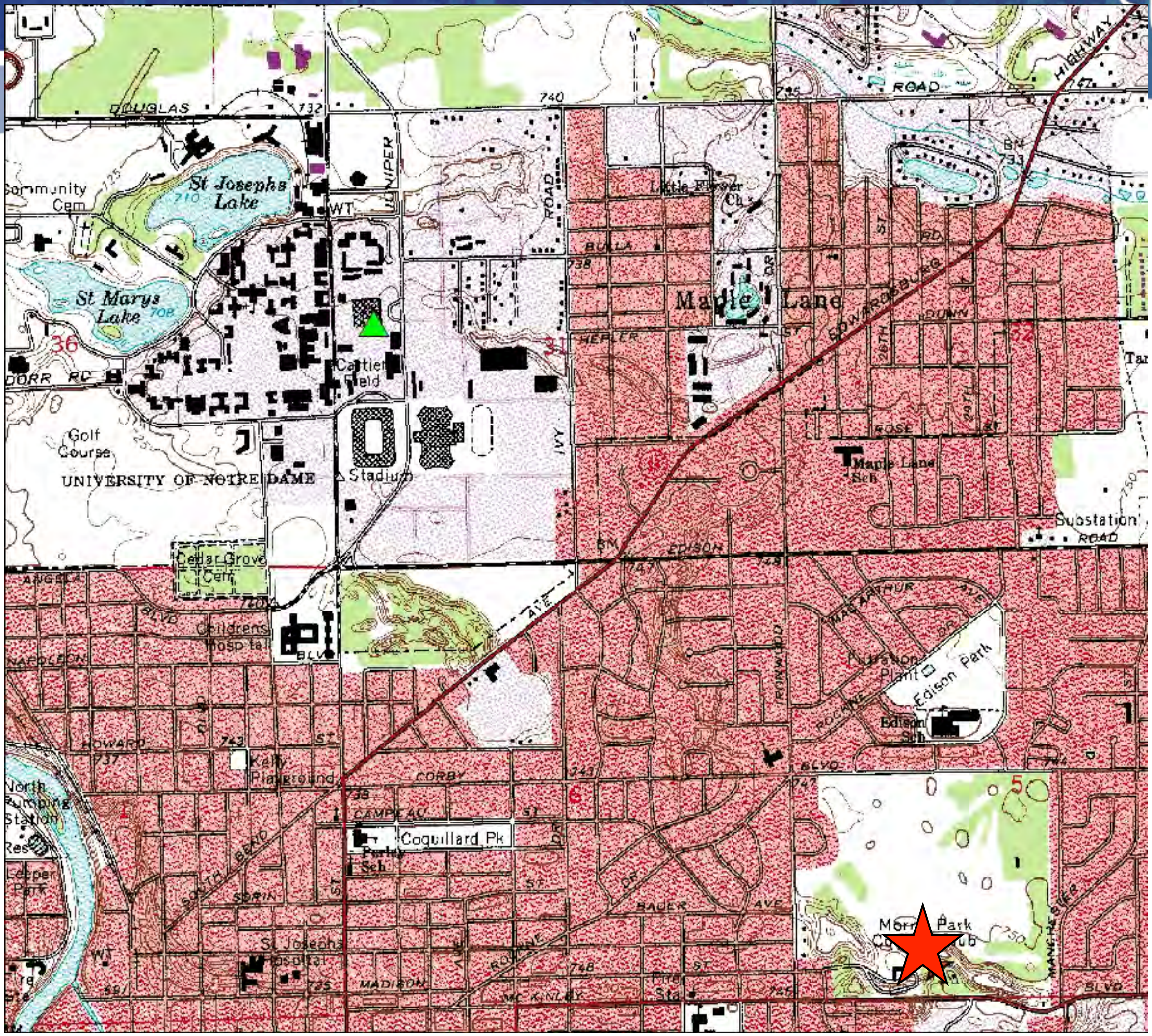
30.01 - 40

40.01 - 50

50.01 - 85

Cost Weighted Distance

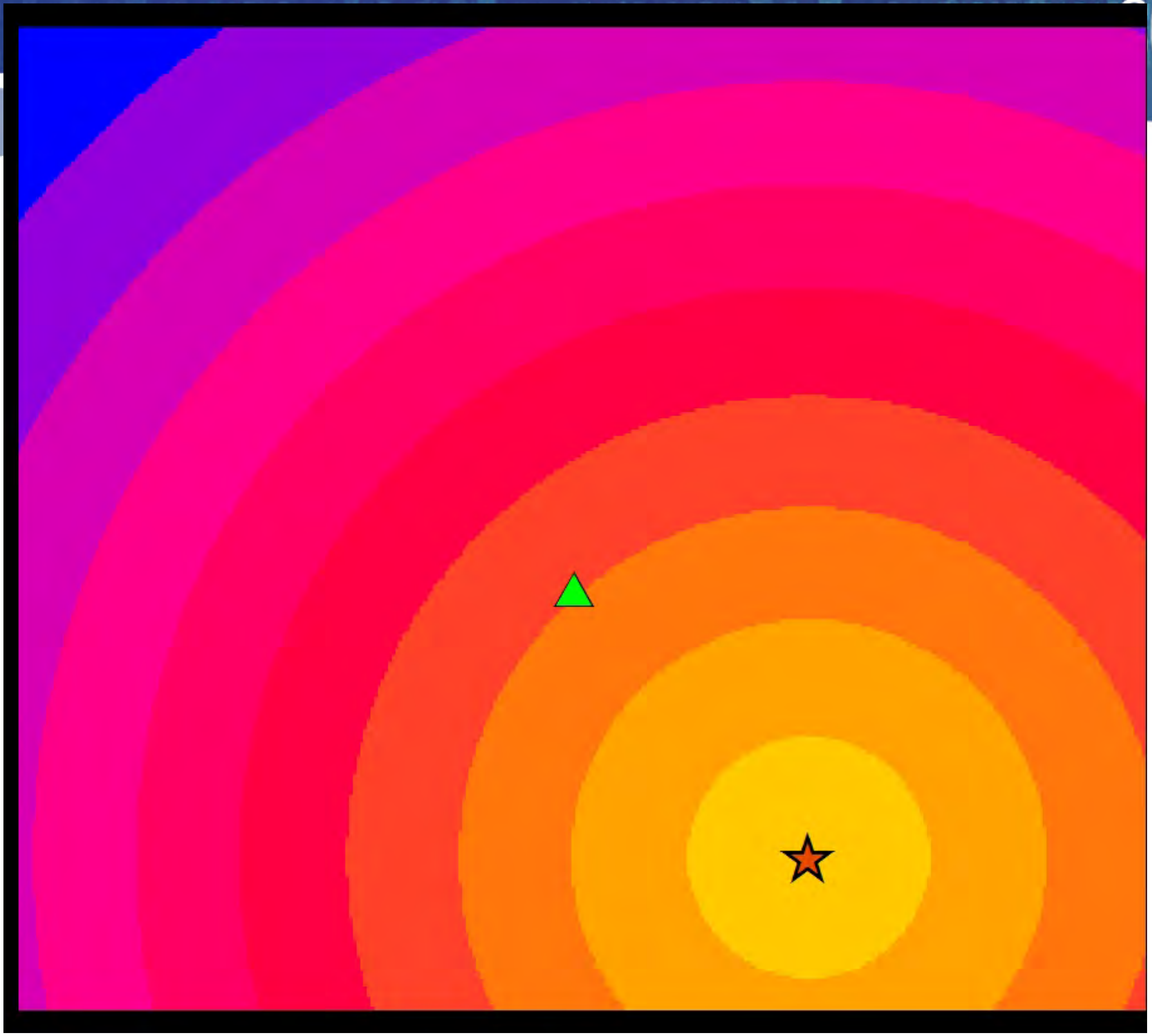
CENTER for DIGITAL SCHOLARSHIP



Say we wanted to walk between the Library and Morris Park Country Club

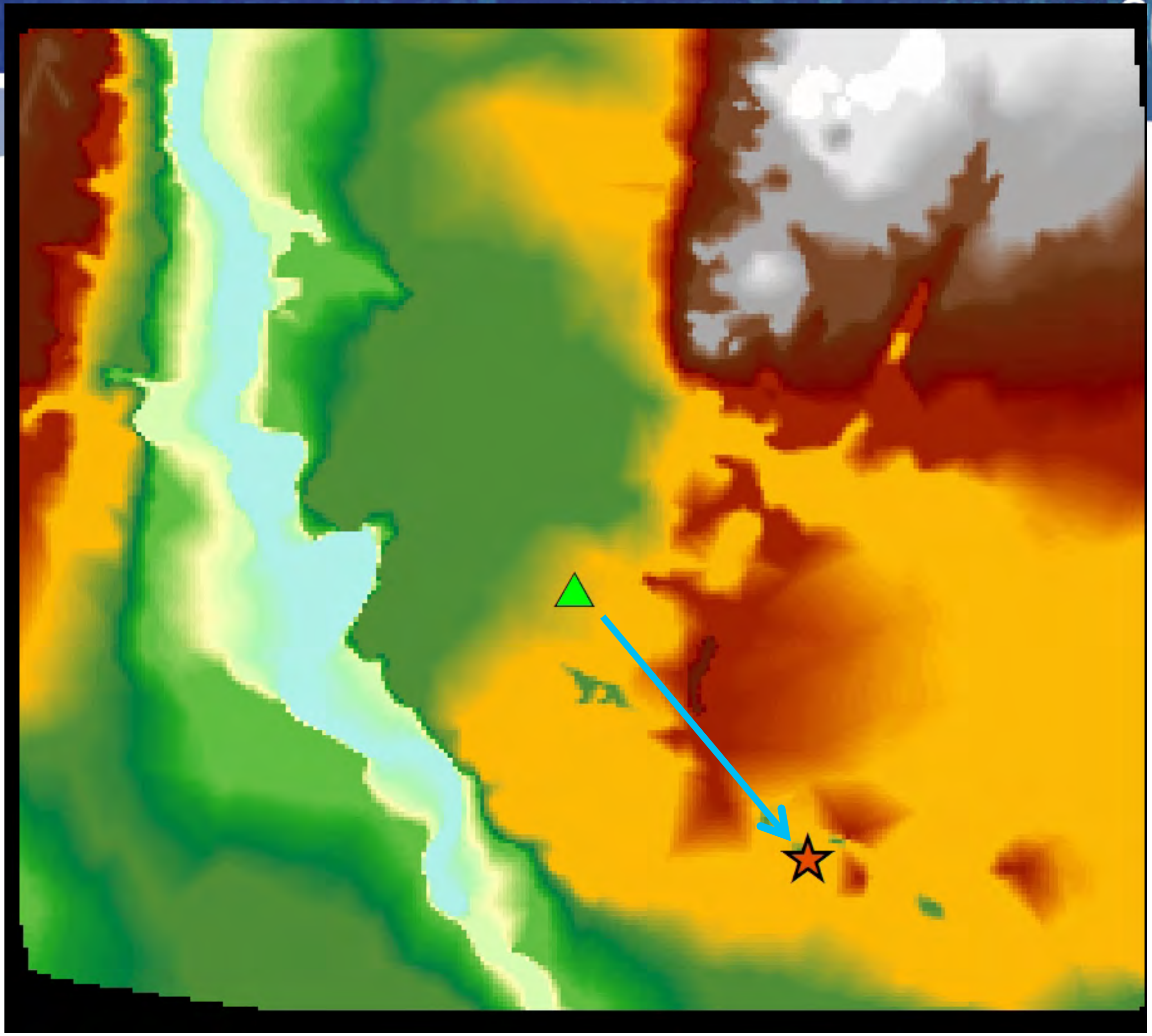
UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Cost Weighted Distance



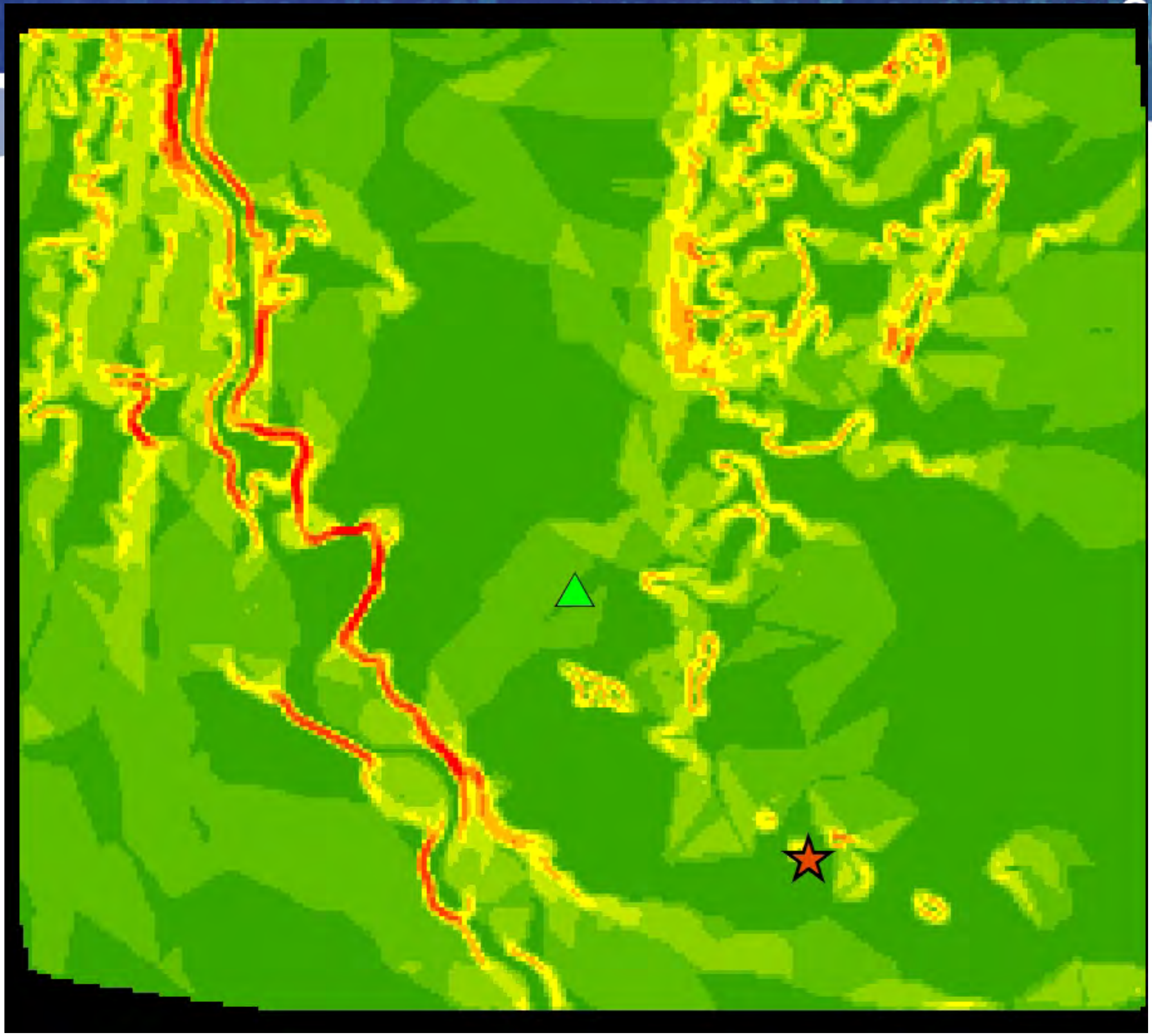
A Euclidean distance calculation would give us a measure of how far away our destination is.

Cost Weighted Distance



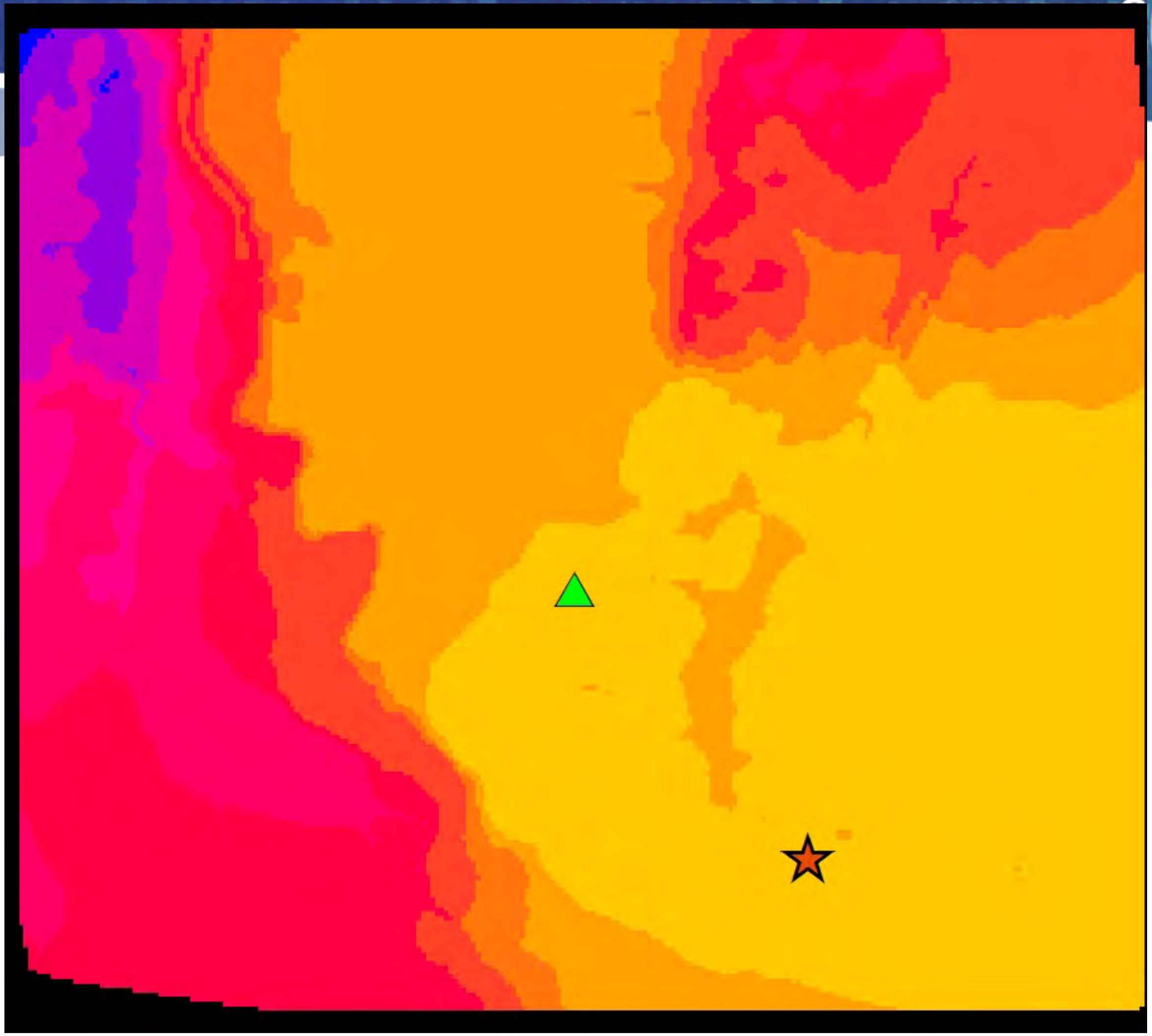
But, if we look at a DEM, we can see that the travel in a straight line would traverse some difficult terrain

Cost Weighted Distance



A slope calculation for this DEM gives us some indication of how easy it would be to travel (higher slopes are harder to cross)

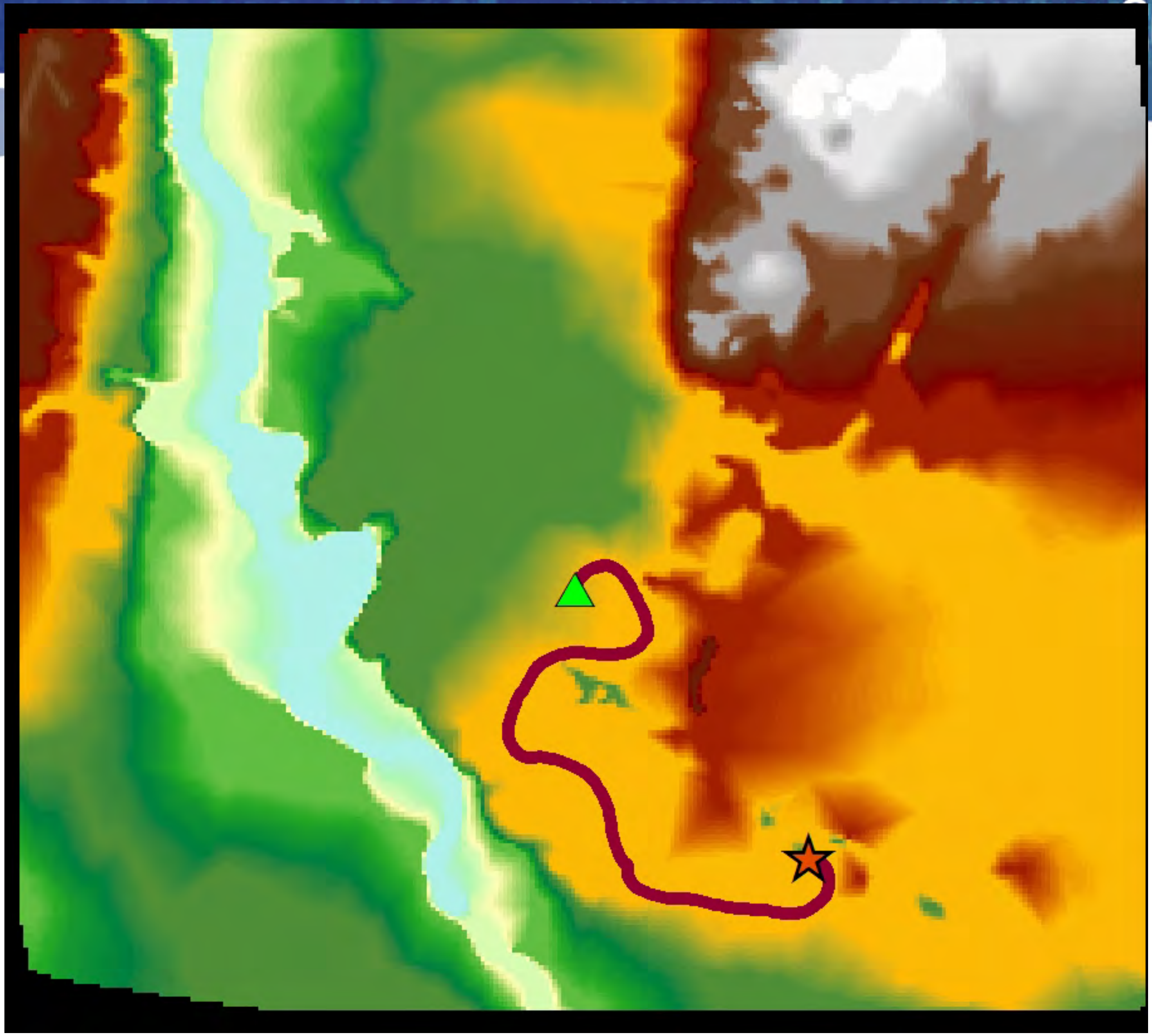
Cost Weighted Distance



Using this slope raster as the “weight” we can create an image that gives us a measure of how difficult it is to travel

-Note that this image shows “ease of travel” not real distance

Cost Weighted Distance



From this cost-weighted distance layer we can calculate the optimal route

Spatial analysis examples

CENTER for DIGITAL SCHOLARSHIP

- **Vector**
  - Interactive selections
  - Spatial joins
- Zonal statistics
- **Raster**
  - Distance and space allocation
  - Cost weighted distance
  - Raster Algebra and complex calculations

UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

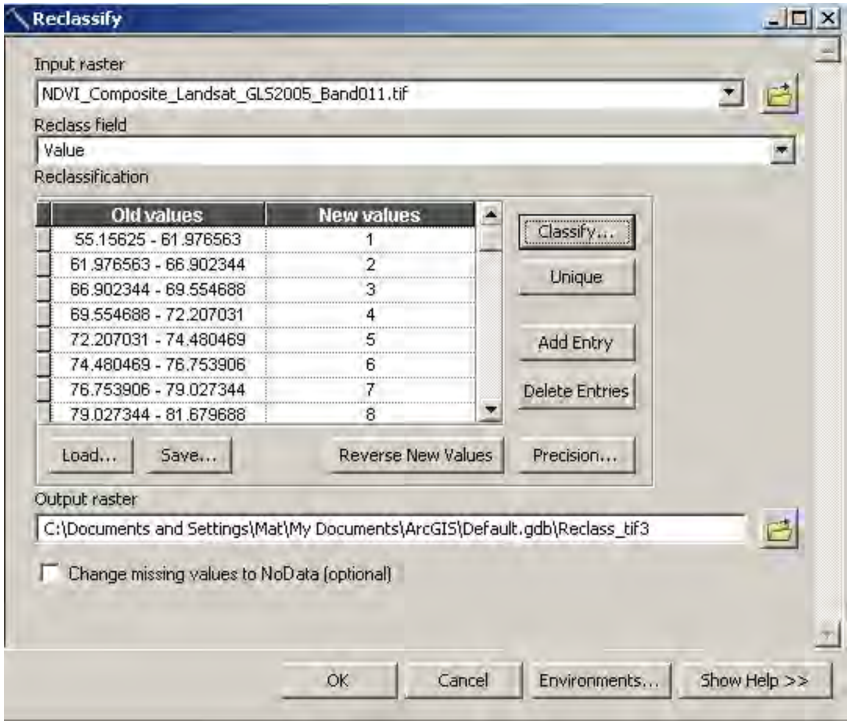
Raster Algebra and complex calculations

CENTER for DIGITAL SCHOLARSHIP

- The process of combining multiple layers to model some characteristic
  - Layers can be weighted according to their importance
- Often, each layer is processed independently to reflect its importance for the final model
  - This process is called reclassifying
- The final model is usually created by mathematically combining these layers

Reclassifying

- Systematically assigning new values to a raster image
  - Based on ranges of old values
- Can be used to
  - Adjust different layers to the same scale
  - Rank different values based on their importance in the final model



Types of reclassifications

Two main types:

- Ranked (1 – 5, 1 – 10, 1-3, etc.):
  - Reducing all of the variability into a series of ranked classes
  - You decide what should be ranked high or low
  - Several very different types of data are then scaled the same and can be combined into a model
- Binary (0 or 1)
  - Weighting a character as either possible (1) or not possible (0)
  - If we then multiply this variable by the rest of the components of the model, it will remove any areas that are not possible
  - Example: No matter how highly ranked the area, we cannot build a new coffee shop in the middle of a lake
    - So we would want areas of water to always be coded as low

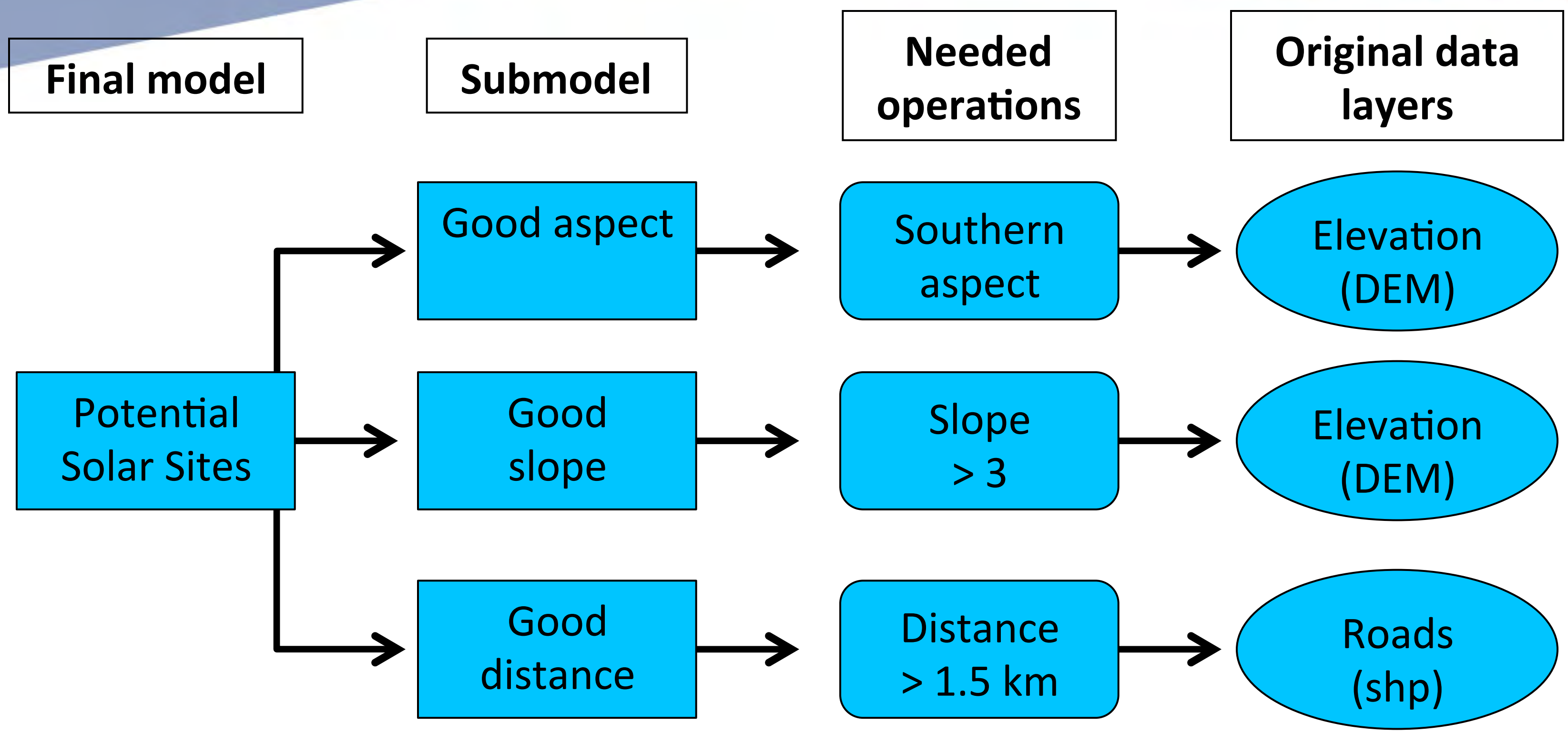
Example

CENTER for DIGITAL SCHOLARSHIP

- We want to find possible locations for solar panels on Notre Dame Campus

A model of good locations for solar panels

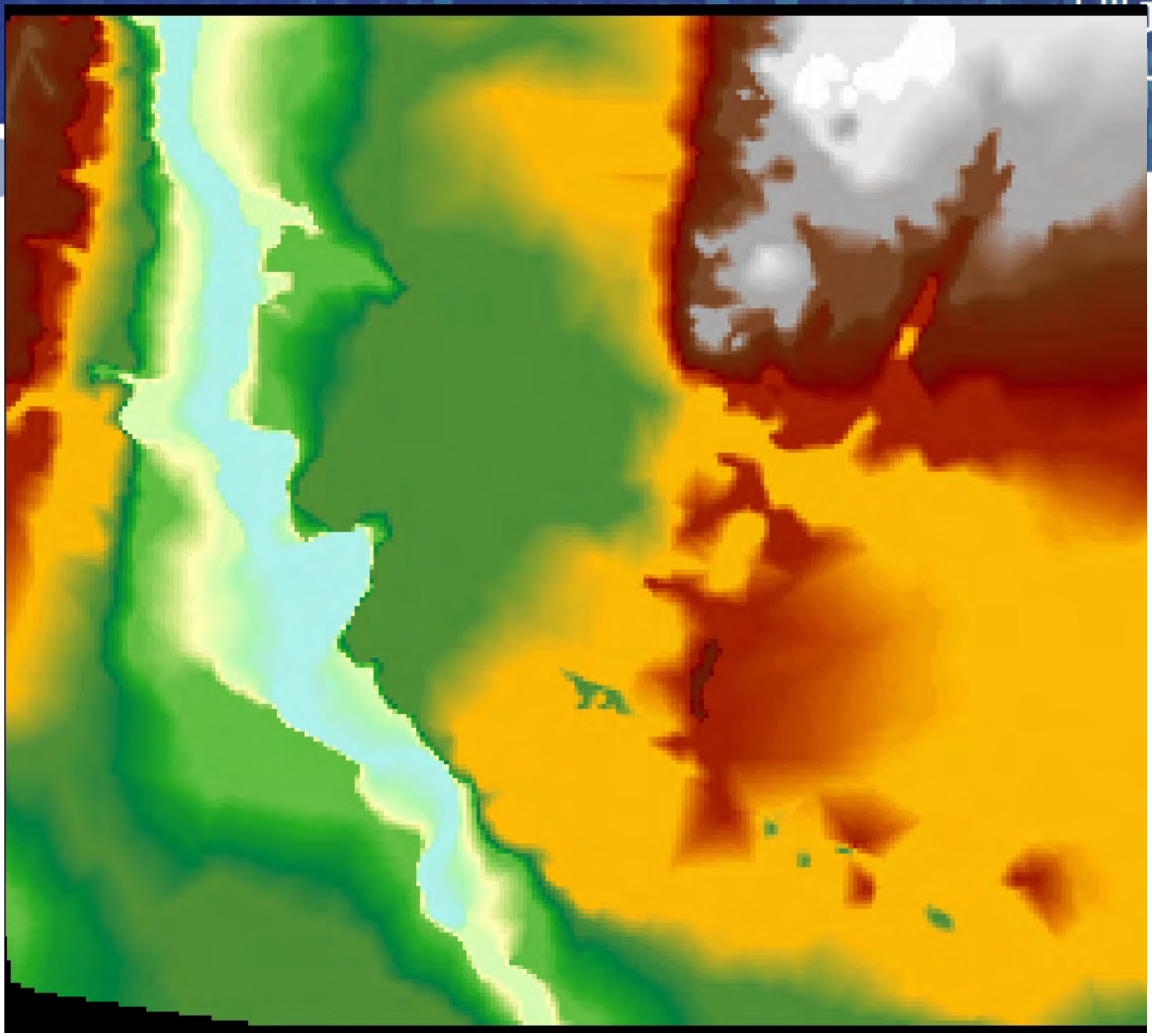
CENTER for DIGITAL SCHOLARSHIP



- We can manipulate data that we have seen to create the components of this model
- Each component should be binary:
  - Either we are far enough from roads or not

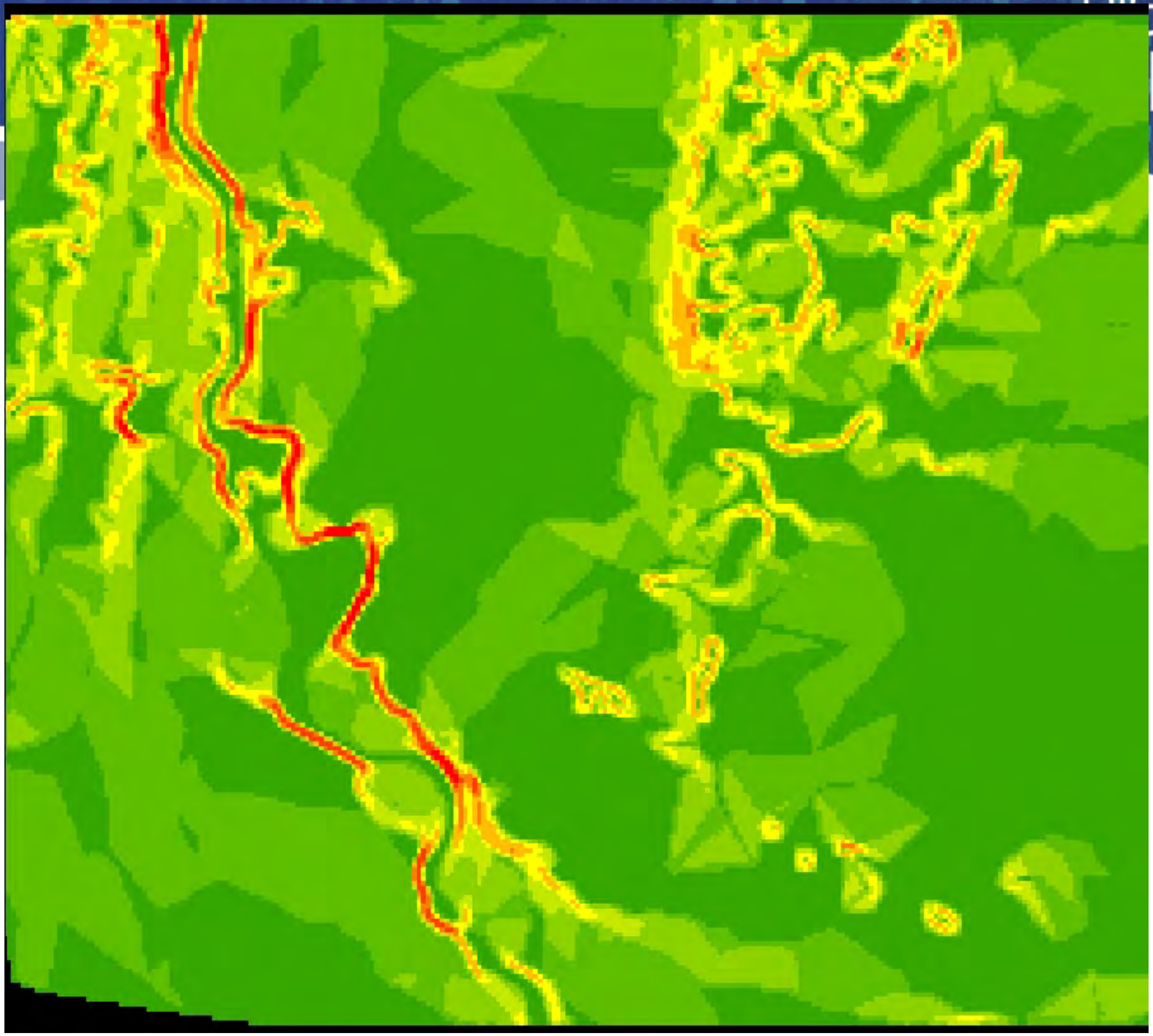
UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Original Data Layer: Elevation



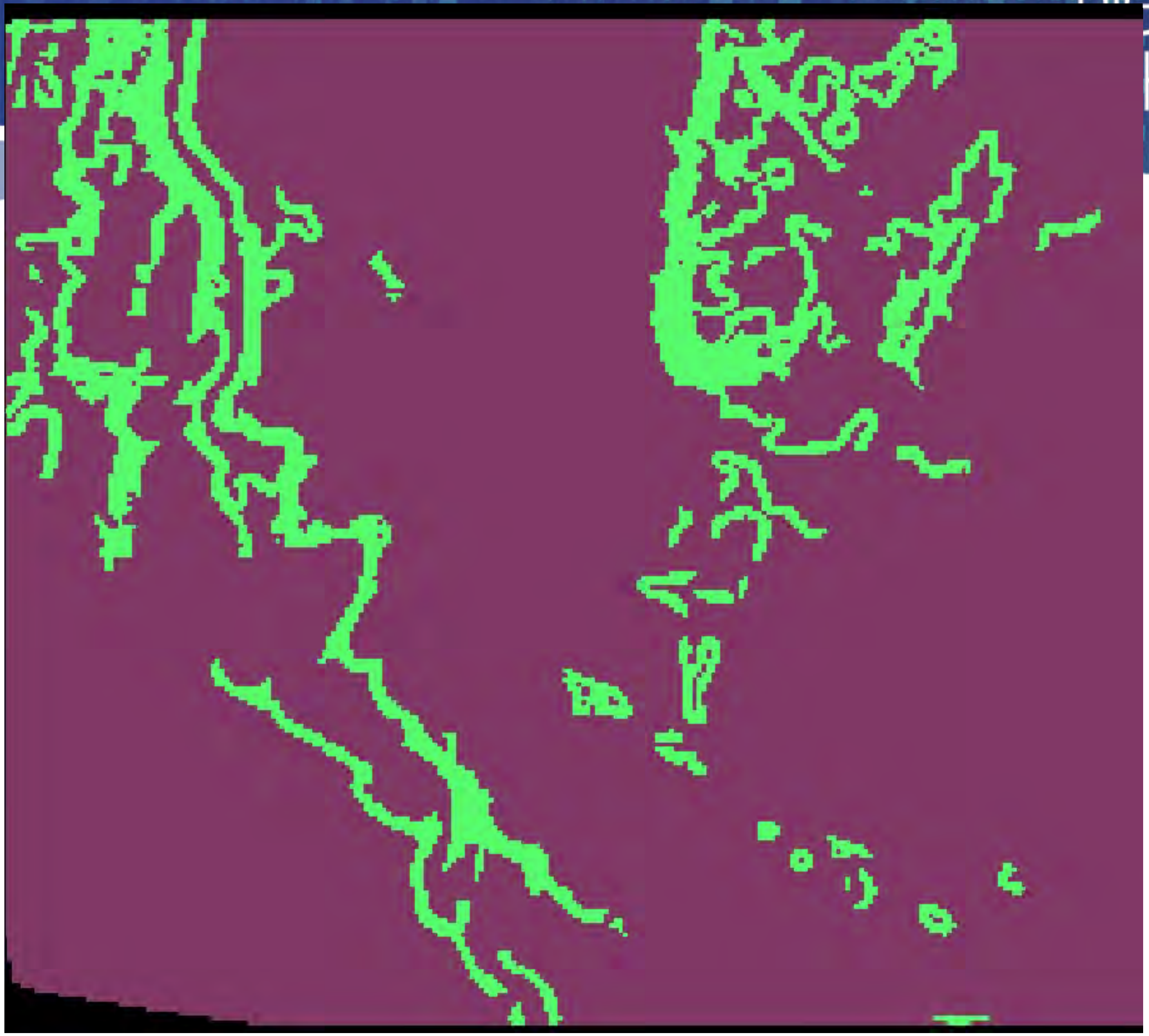
The DEM, showing elevation, is our first data layer to process

Processed Data Layer: Slope



From the DEM we are able to calculate a raster image of the Slope
We can then reclassify this into the image we need

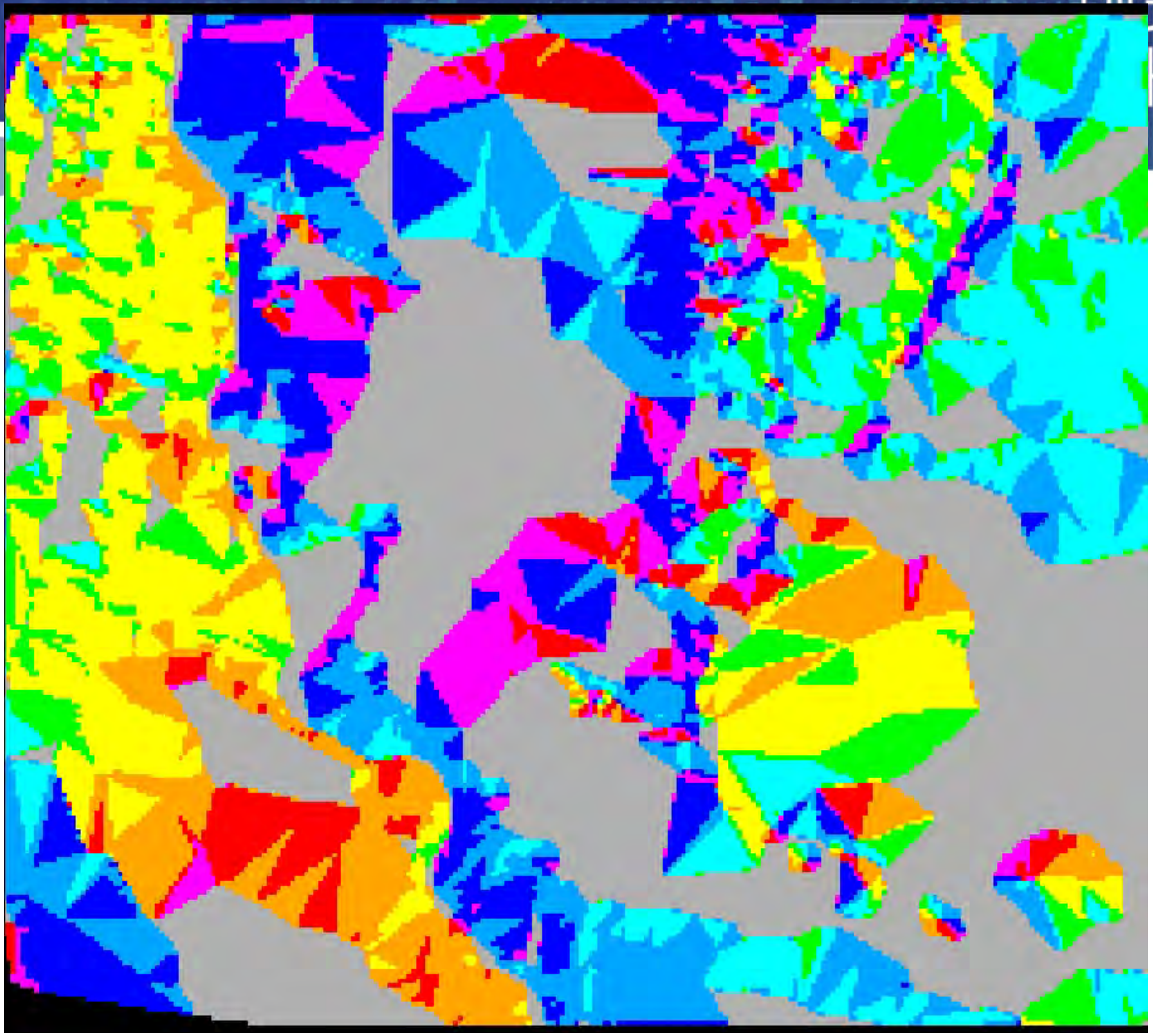
Submodel: Good Slope



Using the operation of slopes greater than 3, we then reclassify the image into good (>3, value 1) and bad (<3, value 0)
Here the green areas are good slopes

Processed Data Layer: Aspect

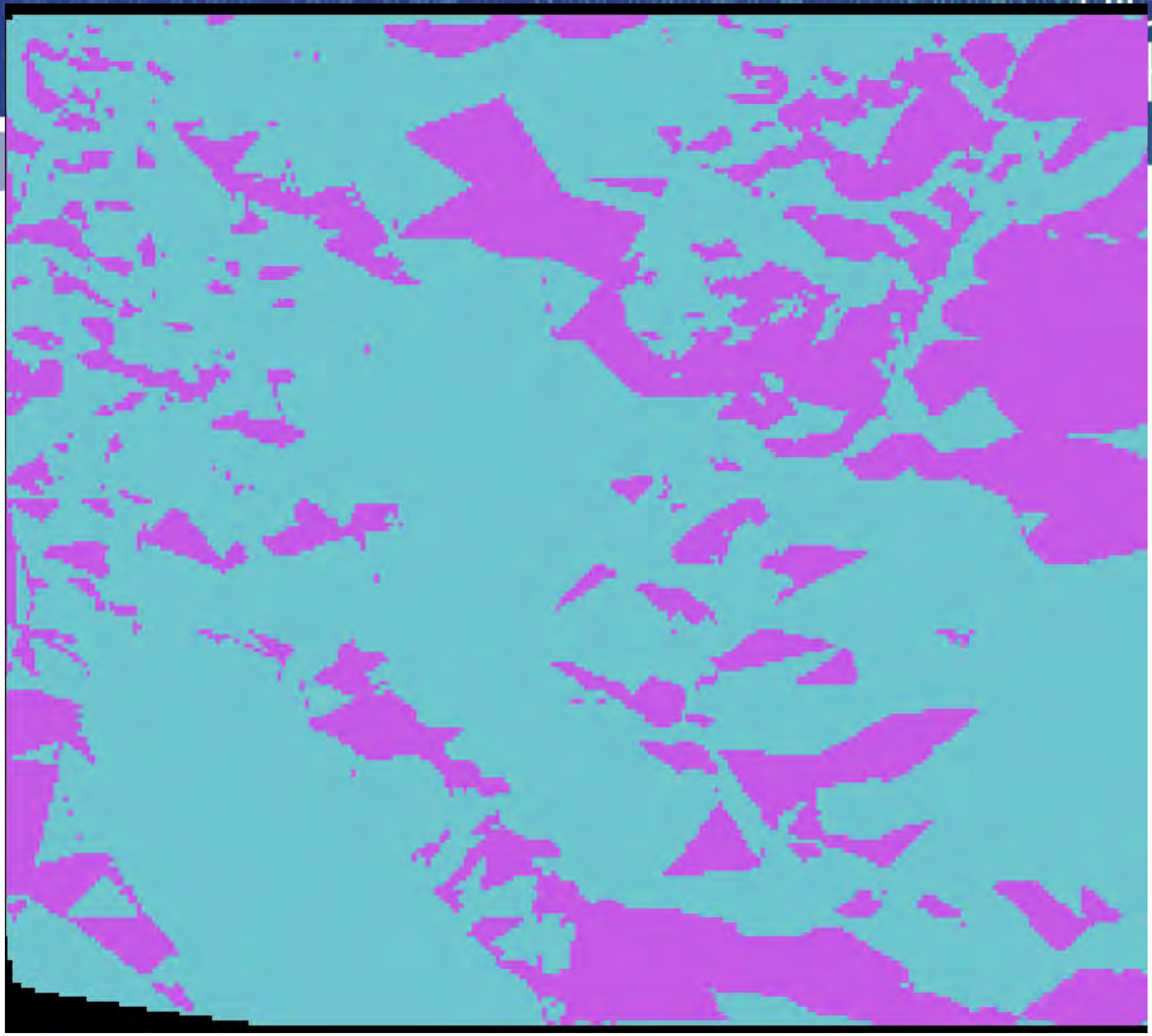
CENTER for DIGITAL SCHOLARSHIP



From the DEM we can also calculate a raster image of the Aspect
We can then reclassify this into the image we need

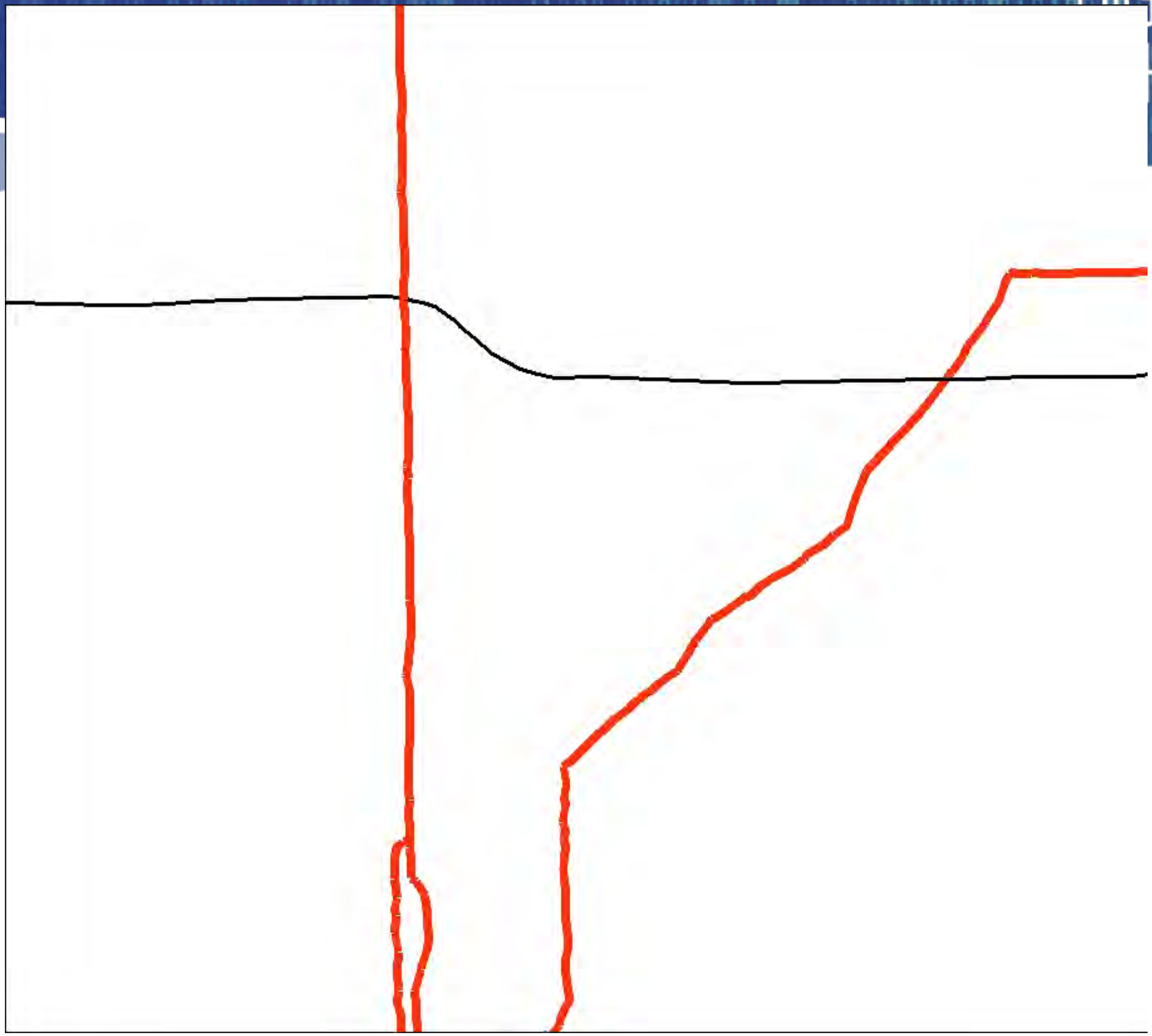
UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

Submodel: Good Aspect



Aspects greater than 3: Reclassify the image into good (south facing, value 1) and bad (not south facing, value 0)
Here the pink areas are good aspects

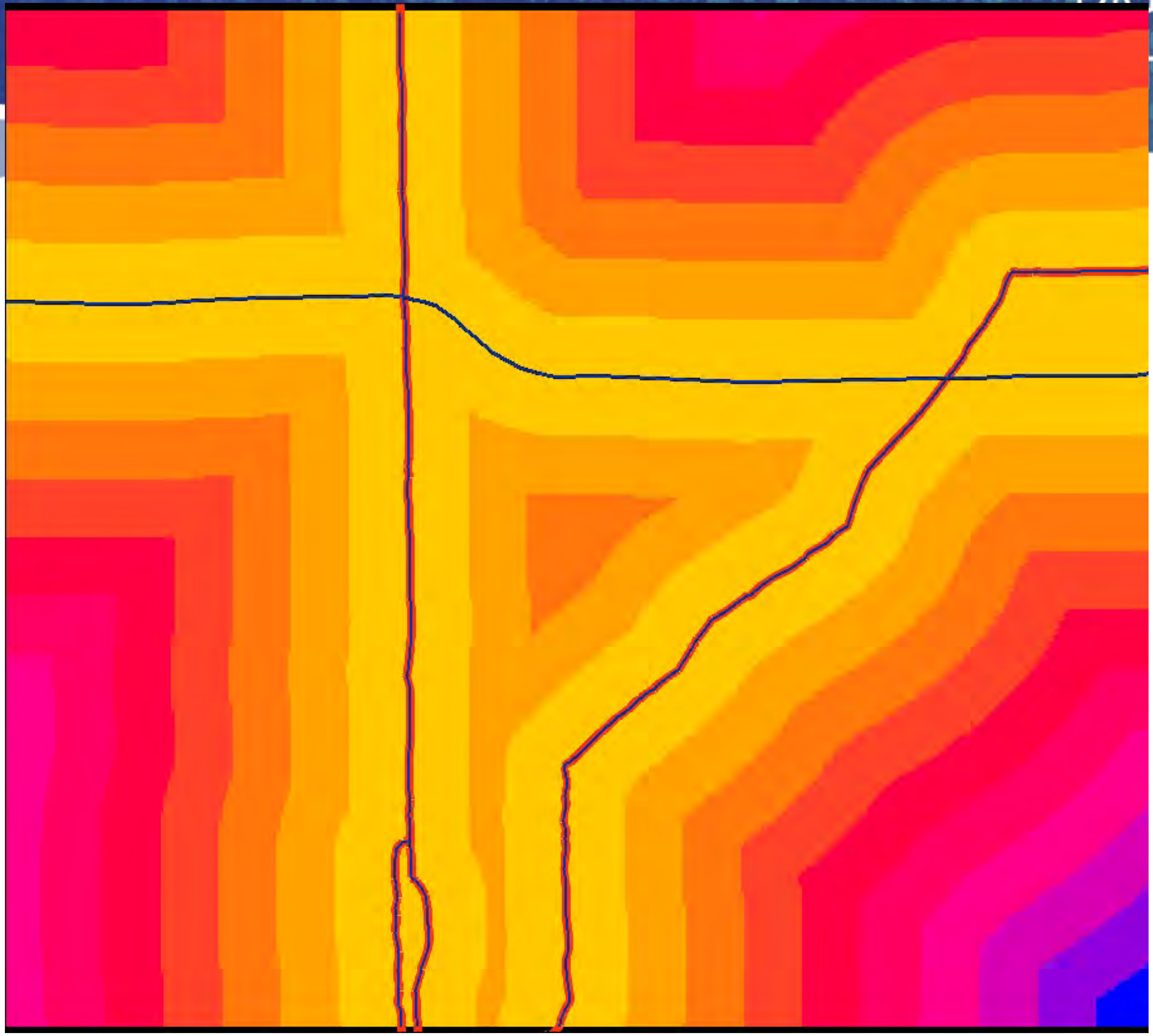
Original Data Layer: Roads



The vector shapefile showing roads serves as a base data layer to calculate distances from the roads

Processed Data Layer: Distance to roads

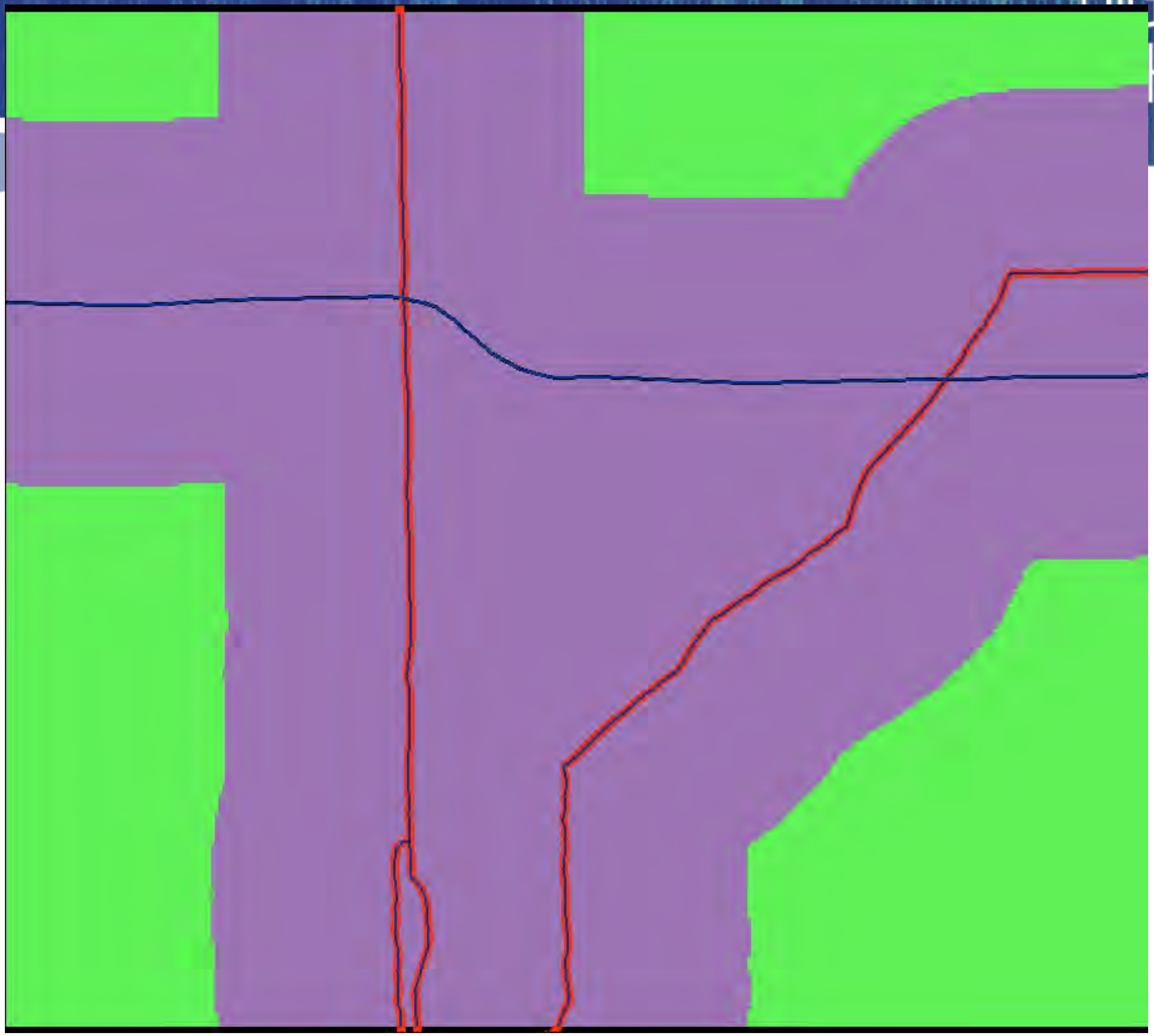
CENTER for DIGITAL SCHOLARSHIP



A Euclidean distance calculation on the roads gives us the distance to the closest from any part of the image

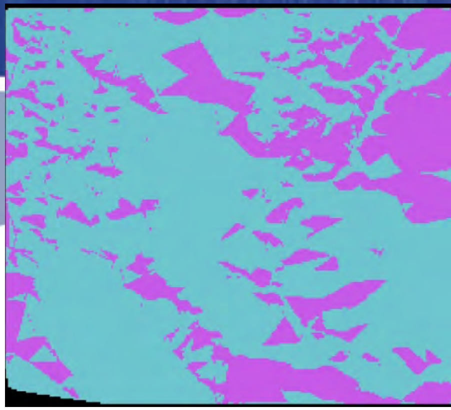
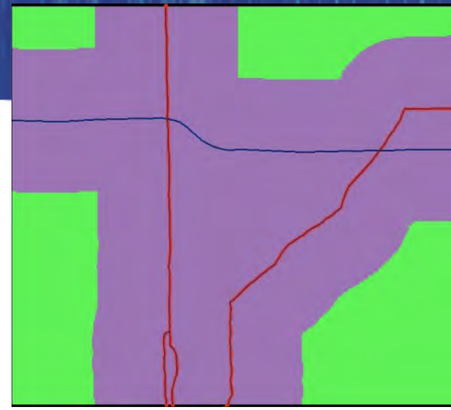
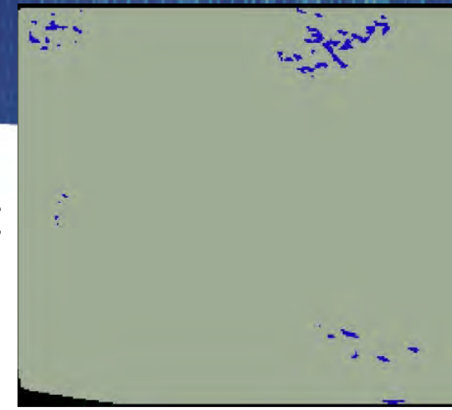
UNIVERSITY OF NOTRE DAME
Hesburgh Libraries

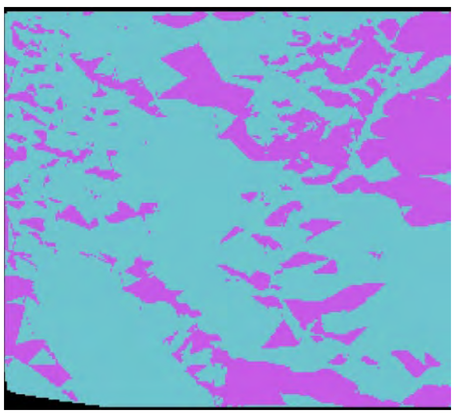
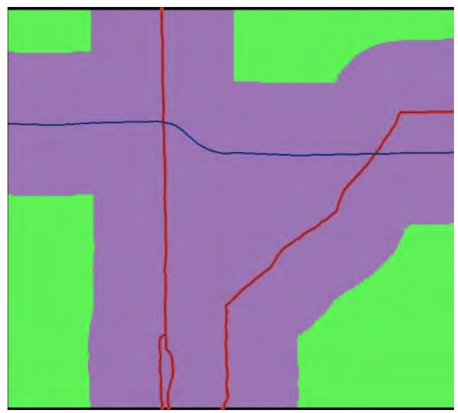
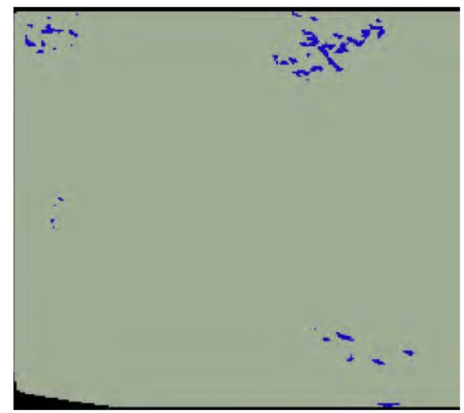
Submodel: Good distance to roads



Using the operation of distances greater than 1.5 km, we then reclassify the image into good (>1.5, value 1) and bad (<1.5, value 0)

Final calculation

 *  *  =

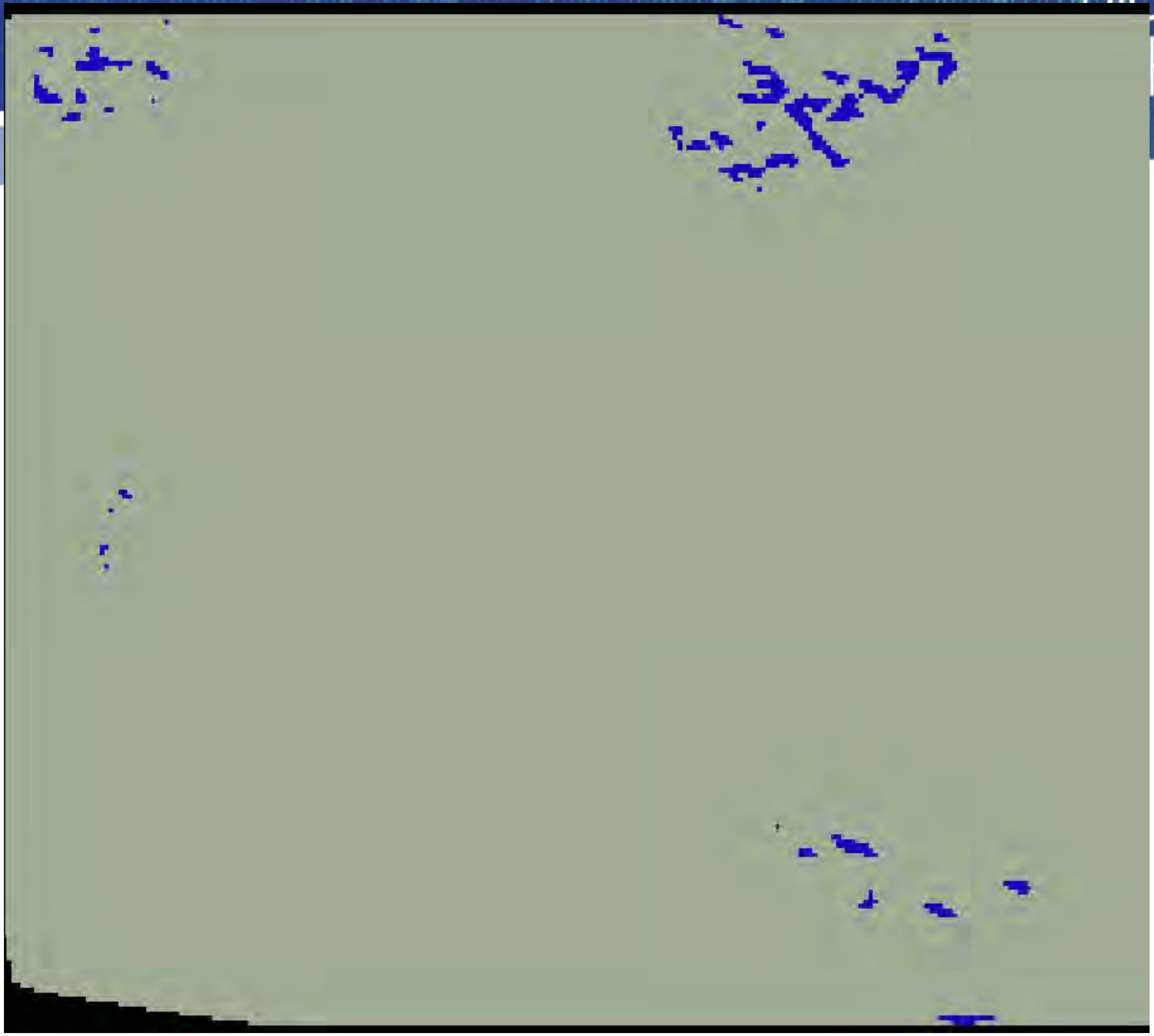
 +  +  =

We now combine the tree layers into the final model of Potential Solar Sites

We choose if we want to multiply the characters to get binary final image

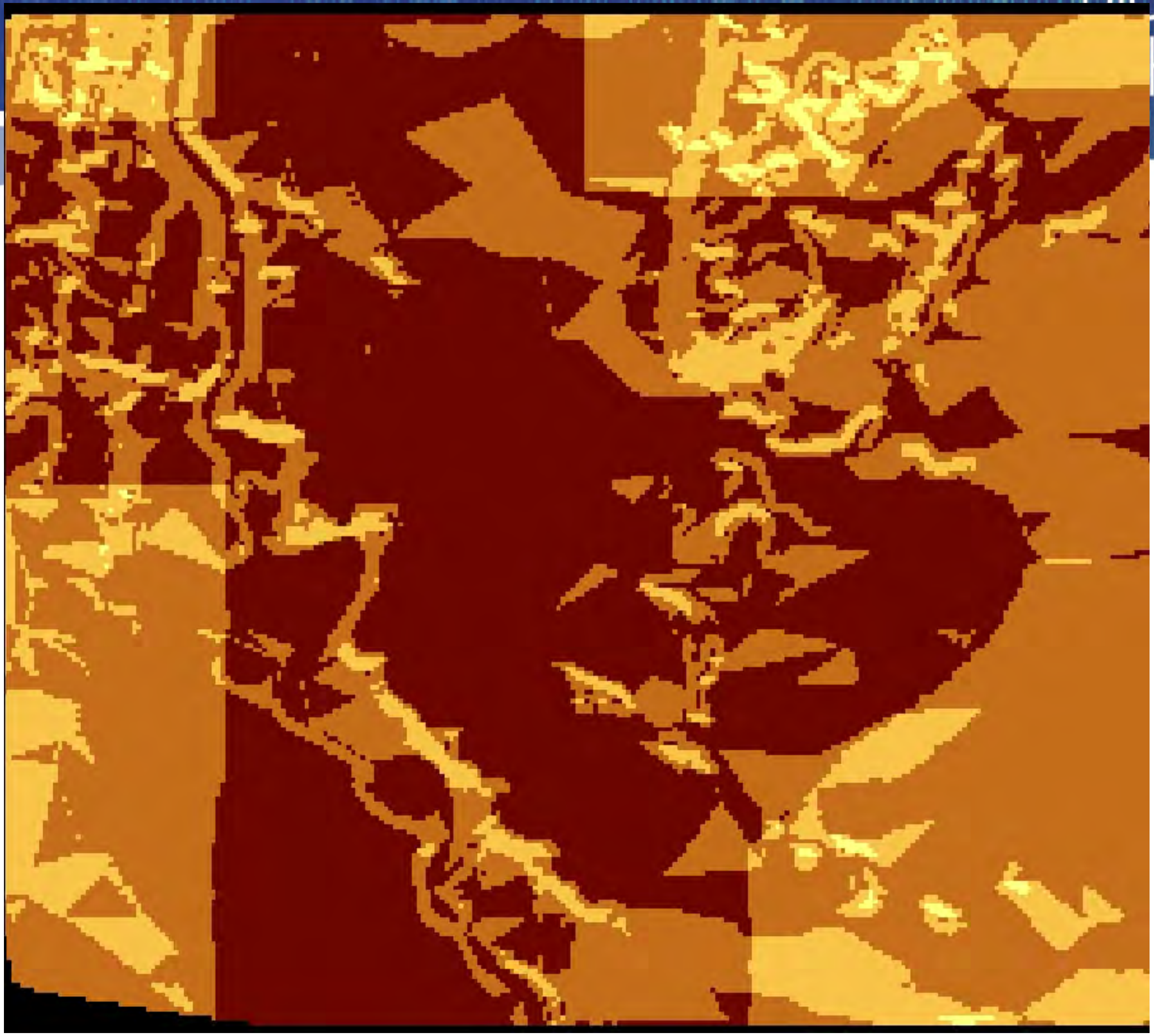
Or add them to get a ranked final image

Final Model: Binary



Each pixel is coded as Unsuitable (0, tan) or Suitable (1, blue)

Final Model: Ranked



Each pixel is coded as 0 – 3 based on the number of criteria it met.
Anything with 3 is highly suitable for a solar panel