

Automation and Flexibility with ArcPy: A Stand-Alone Python Tool for Retrieving and Analyzing Natural Hazard Data

GIS 6389 – Master's Project

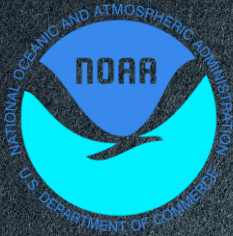
Student: David Lindsey

Advisor: Dr. Fang Qiu

2018 May 4th



What Types of Natural Hazard Data?



NOAA

- Hail (1955-2017)
- Hurricane (1842-2017)
- Tornado (1950-2017)
- Wind (1955-2017)



USGS

- Earthquake (1900-Present)



What is the Tool's Purpose?

- Provide researchers with a simple application to quickly retrieve and analyze various natural hazard types.
- Improve time-management through automation.
 - *Preparing output workspaces*
 - *Finding the web URLs*
 - *Retrieving CSVs & Shapefiles*
 - *Checking data for errors*
 - *Extracting only the necessary data*
 - *Converting CSVs*
 - *Assigning Coordinate Systems*
 - *Clipping data to area of interest*
 - *Performing analysis on data*
 - *Extracting count totals and summary statistics*
 - *Designating output naming conventions*



AUTOMATED

What is the Tool's Purpose?

Quickly answer questions concerning the data.

- *How many tornadoes occurred in Texas from 1970 to 2017? Which counties received them, how many per county, and what was the maximum tornado size recorded in each county?*
- *How many hurricanes struck the United States in 2005, and what were the names of those storms? Which states were hit by those storms, and what were the wind speeds?*

What is the Tool's Purpose?

- Complex tools such as FEMA's "Hazard US" (HAZUS) exist. However...
 - *Application size by itself is 4+ GB.*
 - *Complete dataset for entire United States is 60+ GB.*
 - *HAZUS versions are limited to specific versions of ArcGIS Desktop.*
 - *Version 4.0 = ArcGIS 10.4, Version 4.2 = ArcGIS 10.5, etc.*
 - *Only focuses on earthquakes, hurricanes, floods, and tsunamis.*
 - *Currently offering 7 training courses for new users.*
 - **Not a practical application for most GIS users.**



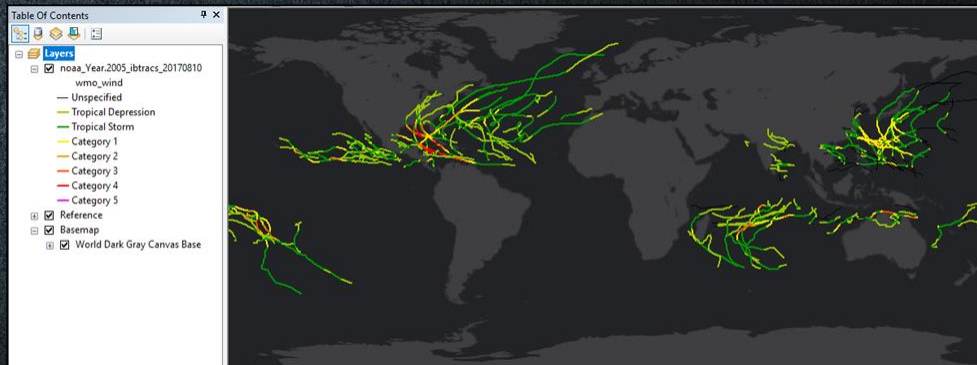
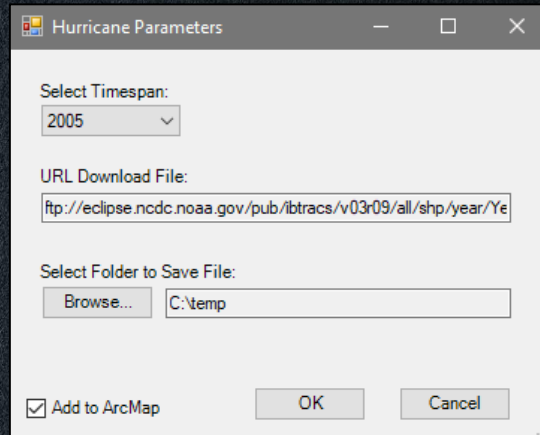
<https://www.fema.gov/hazus>

Background

Inspired by a previous, small project I created:

- VB.NET/ArcObjects Add-In Tool for ArcMap:
 - Retrieved, displayed, and symbolized hazard data within ArcMap.
 - **It did nothing more than this.**

The long-term goal was to redesign, expand, and improve this concept.



Implementation and Reasoning

Stand-alone application created with ArcGIS Pro's Conda-based Python 3.x.

- *ArcGIS Pro and Python 3.x are now mainstream.*
- *Python 2.x support is scheduled to end in January 2020.*
- *“Forward-Thinking” without alienating ArcGIS Desktop users.*
- *Users can run this application, then choose to open ArcMap or ArcGIS Pro to load output results.*

Utilizes no third-party modules (other than ArcPy) in order to function.

- *E.g. PyQt, WxPython, among others.*
- *As naturally “Pythonic” as possible.*
- *Users are able to launch the application without needing to install other Python prerequisites besides ArcGIS Pro.*

Prerequisites and Assumptions

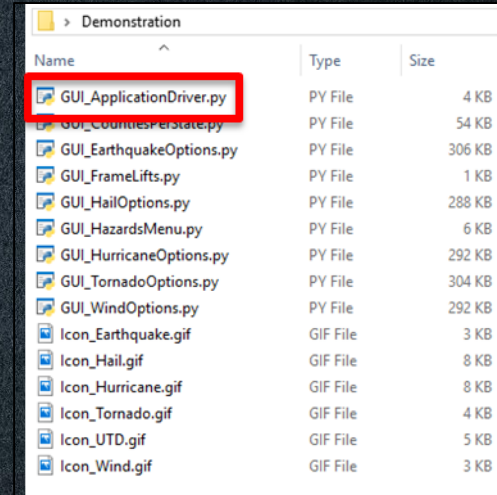
- The user already uses ArcGIS, with ArcGIS Pro 1.4 or higher installed.
 - *ArcPy geoprocessing tasks require active Esri license.*
 - *Minimum of Basic License for ArcPy functionality.*
 - *Advanced License preferred, due to Spatial Analyst analysis options.*
- User has access to stable internet connection.
 - *Application relies on downloadable data.*
- User computer possesses Windows 8.1 or Windows 10.
 - *Testing has been unsuccessful with Windows 7 Enterprise.*
 - *ArcGIS Pro does not support anything less than Windows 7 Professional.*
 - *Microsoft support for Windows 7 ending January 2020.*

Steps to Launch Application

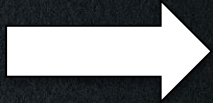
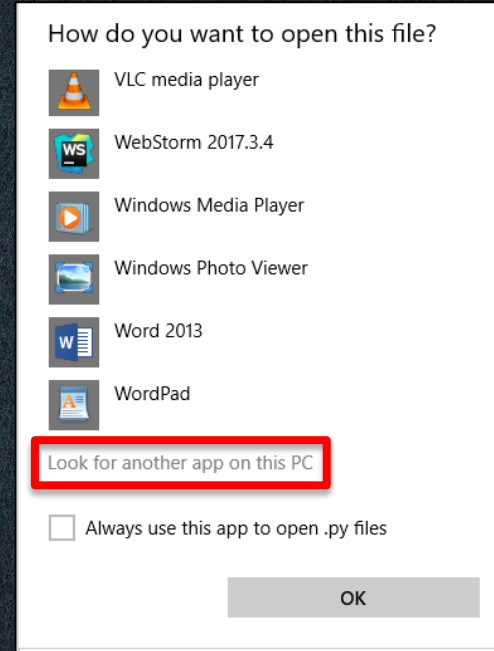
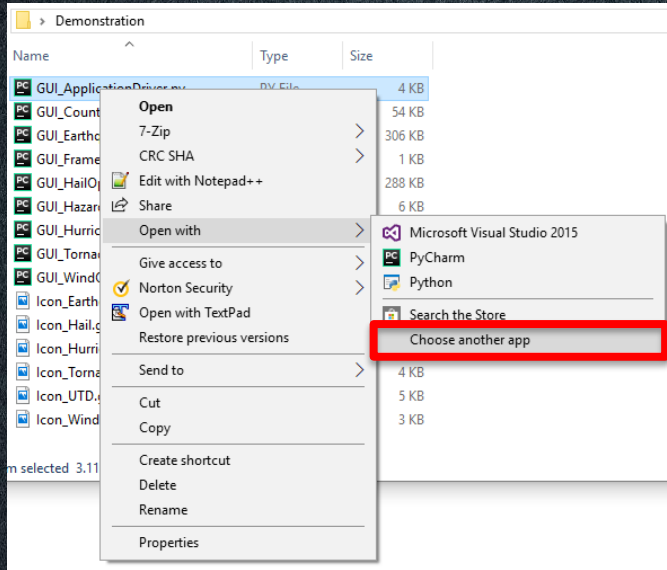
Ensure you are:

- a) currently signed into ArcGIS Pro,
- or-
- b) have ArcGIS Pro authorized to work off-line.
- ArcGIS Pro window does not need to be open for this tool to work.

Then navigate to the launch file called “GUI_ApplicationDriver.py”



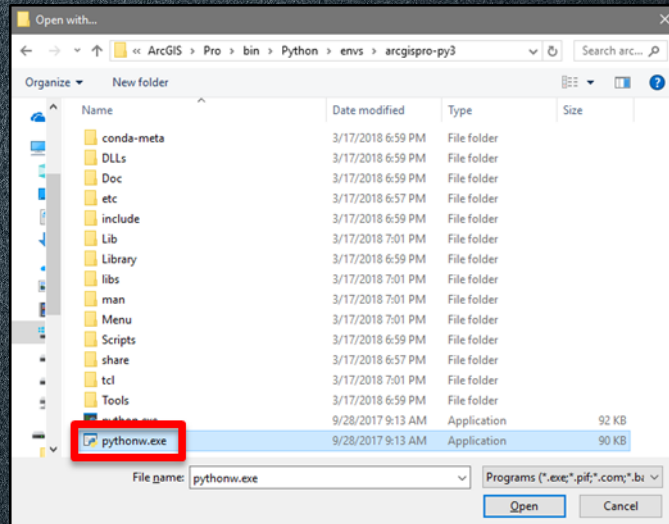
Launching Application (cont.)



Launching Application (cont.)

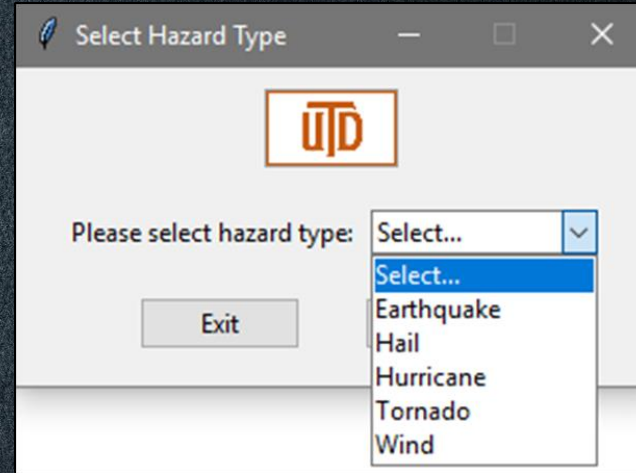
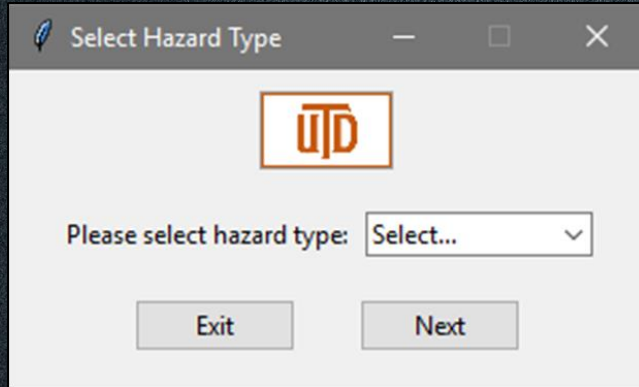
Navigate to:

- C:\Program Files\ArcGIS\Pro\bin\Python\envs\arcgispro-py3\pythonw.exe
- Click Open.



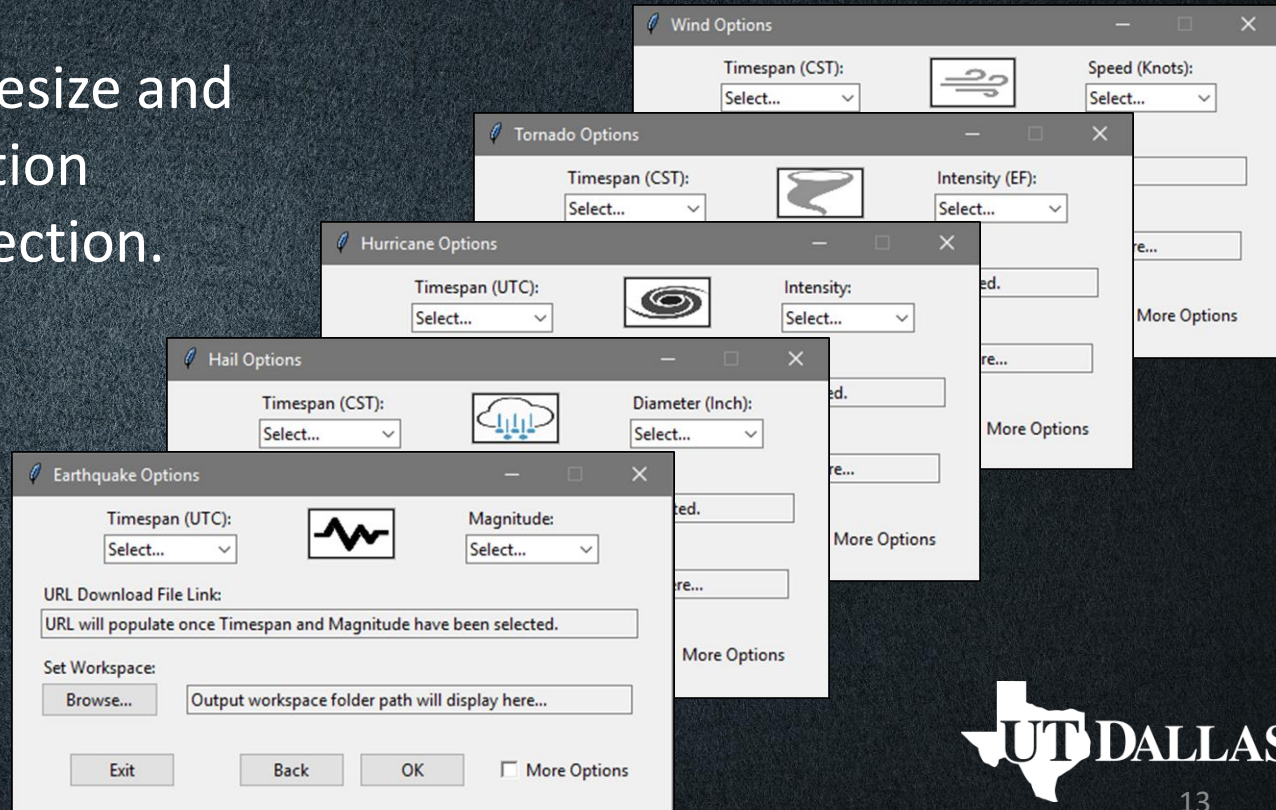
Note: C:\Program Files\ArcGIS\Pro\bin\Python\Scripts\propy.bat will work as well, but it will launch the Windows Command Prompt in tandem.

The application has launched!



Natural Hazard Options

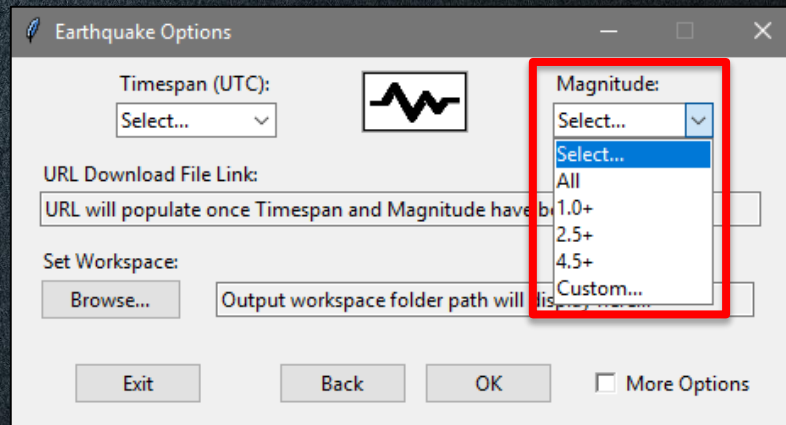
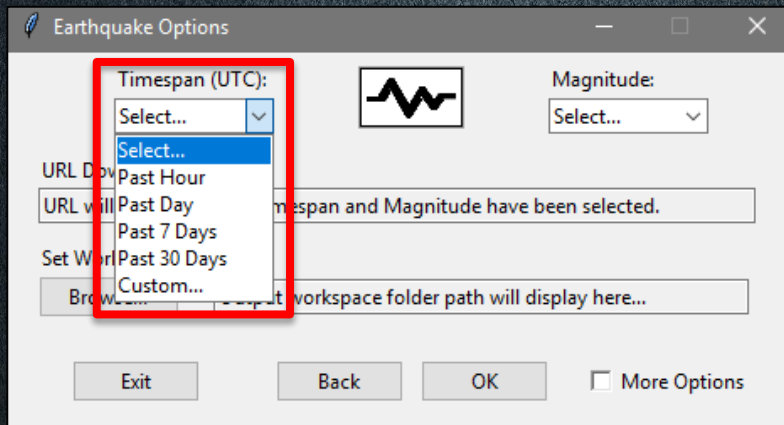
GUI window will resize and display hazard option based on user selection.



Hazard Parameters

Earthquake Example:

- USGS **Timespan** (UTC) choices are:
 - Past Hour, Past Day, Past 7 Days, Past 30 Days, and Custom From/To Range with Month/Year (1900 to Present)
- USGS **Magnitude** (Moment Magnitude Scale) choices are:
 - All, 1.0+, 2.5+, 4.5+, and Custom Min/Max Range (-1.0 to 10.0)



Custom Hazard Parameters

Earthquake Example:

- The GUI window will resize and display Custom Parameters options.

Earthquake Options

Timespan (UTC): Custom...

Magnitude: Custom...

From: Year: 2017 Month: 1 **To:** Year: 2017 Month: 12 **Range:** Min: 3.0 Max: 9.0

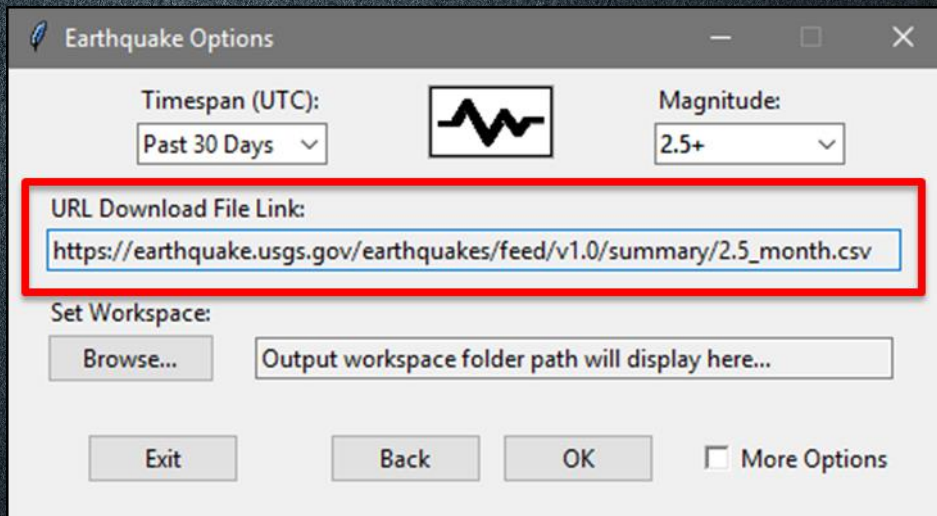
URL Download File Link:
<https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=201>

Set Workspace:
Browse... Output workspace folder path will display here...

Exit Back OK ☐ More Options

Hazard Data – URL Display

Selecting the parameters from the drop-down lists will populate the complete URL for the data to be retrieved. This allows the user the ability to review the full URL path.



Earthquake Options

Timespan (UTC):
Past 30 Days

Magnitude:
2.5+

URL Download File Link:
https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/2.5_month.csv

Set Workspace:
Browse... Output workspace folder path will display here...

Exit Back OK ☐ More Options

Hazard Parameters

Each hazard type has a partially unique, **static** web URL where all data is accessed from. Certain **parameters** can be changed, via the timeframe and severity drop-down lists.

Examples for Earthquakes:

- Magnitude **2.5+** during **Past 30 Days**:

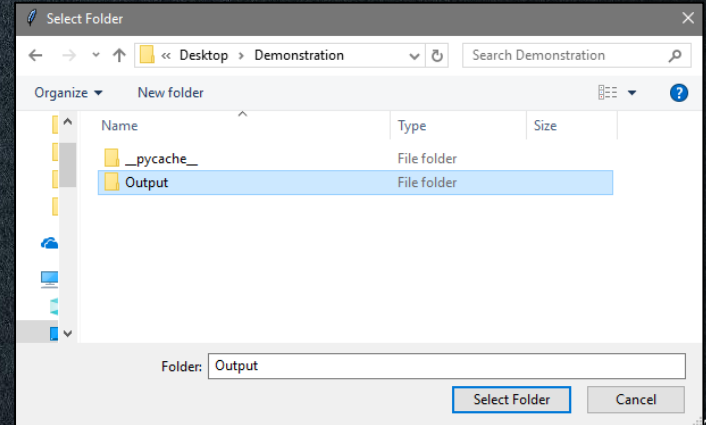
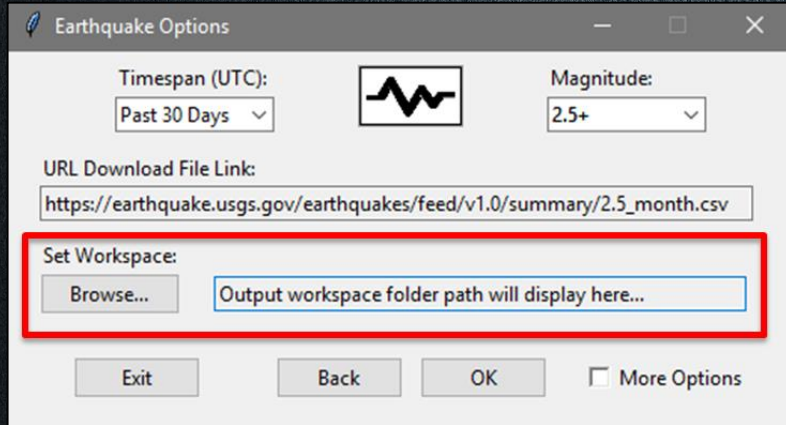
https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/2.5_month.csv

- Magnitude **2.5+** during **Custom Timespan (2017/01 to 2017/12)**:

<https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2017-1-01&endtime=2017-12-31&minmagnitude=2.5>

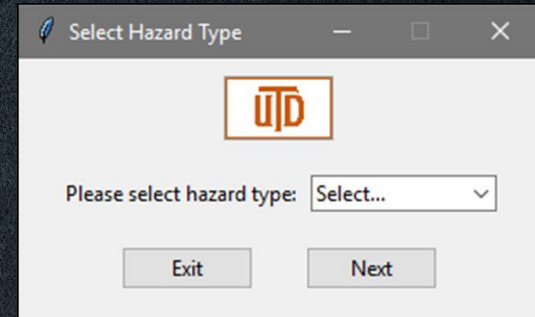
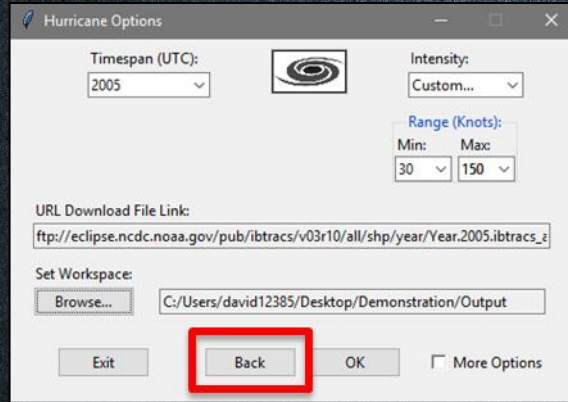
Set Workspace

Selecting the **Browse** button will allow the user to designate the output workspace folder location, where all downloaded/output data will be saved.



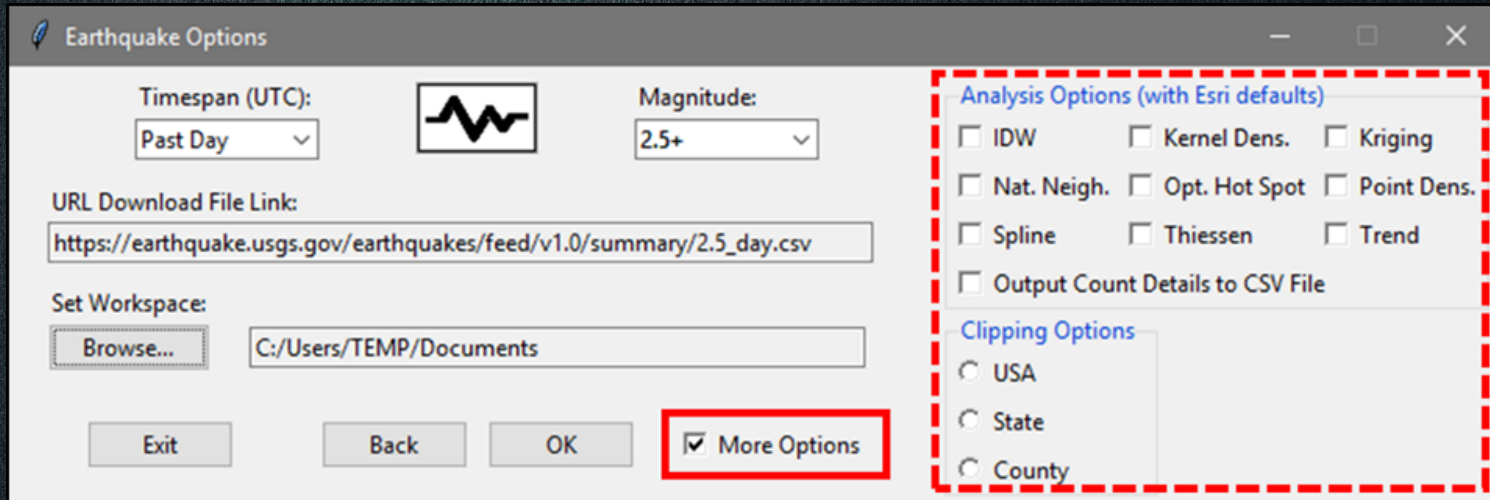
Exit and Back Buttons

- Self-explanatory.
- **Exit** will exit/close the application.
- **Back** will return the user to the previous screen.




More Options

Selecting the **More Options** checkbox will expand the GUI to reveal Analysis Options and Clipping Options.



The screenshot shows the 'Earthquake Options' dialog box. The 'More Options' checkbox is checked and highlighted with a red box. To its right, a dashed red box highlights the expanded 'Analysis Options (with Esri defaults)' and 'Clipping Options' sections.

Earthquake Options

Timespan (UTC):  Magnitude:

URL Download File Link:

Set Workspace:

☒ **More Options**

Analysis Options (with Esri defaults)

- ☐ IDW ☐ Kernel Dens. ☐ Kriging
- ☐ Nat. Neigh. ☐ Opt. Hot Spot ☐ Point Dens.
- ☐ Spline ☐ Thiessen ☐ Trend
- ☐ Output Count Details to CSV File

Clipping Options

- ☐ USA
- ☐ State
- ☐ County

Clipping Options

Three options:

- United States (50 States and Washington, D.C.)
- State
- County/Parish
- Utilizes 2017 U.S. Census Bureau Shapefile

Clipping Options

☒ USA

☐ State

☐ County

Clipping Options

☐ USA

☒ State

☐ County

Clipping Options

☐ USA

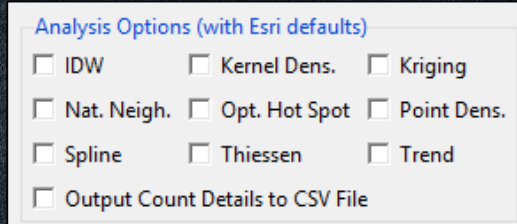
☐ State

☒ County

Within Python, county names are stored within tuples, then assigned to states in a dictionary.

Analysis Options

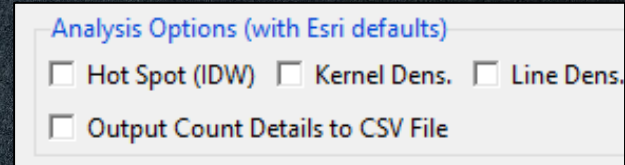
Point-based Options:



Analysis Options (with Esri defaults)

<input type="checkbox"/> IDW	<input type="checkbox"/> Kernel Dens.	<input type="checkbox"/> Kriging
<input type="checkbox"/> Nat. Neigh.	<input type="checkbox"/> Opt. Hot Spot	<input type="checkbox"/> Point Dens.
<input type="checkbox"/> Spline	<input type="checkbox"/> Thiessen	<input type="checkbox"/> Trend
<input type="checkbox"/> Output Count Details to CSV File		

Polyline-based Options:



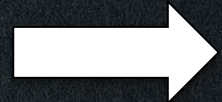
Analysis Options (with Esri defaults)

<input type="checkbox"/> Hot Spot (IDW)	<input type="checkbox"/> Kernel Dens.	<input type="checkbox"/> Line Dens.
<input type="checkbox"/> Output Count Details to CSV File		

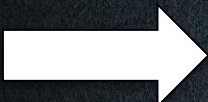
- These are ArcGIS geoprocessing tools
 - (except Output Count Details to CSV File)
- Esri defaults applied to any “optional” parameters.
- If a “magnitude” field is required, it will be added automatically (e.g. hail size, wind speed, etc.)

Output Count Details to CSV File

- Custom operation created specifically for this application.
- Outputs hazard count occurrences and summary statistics based on user's timespan/severity/clipping selections.
- Example:
 - User selects earthquakes
 - Past 30 days
 - 2.5+ Magnitude
 - Clipped to California

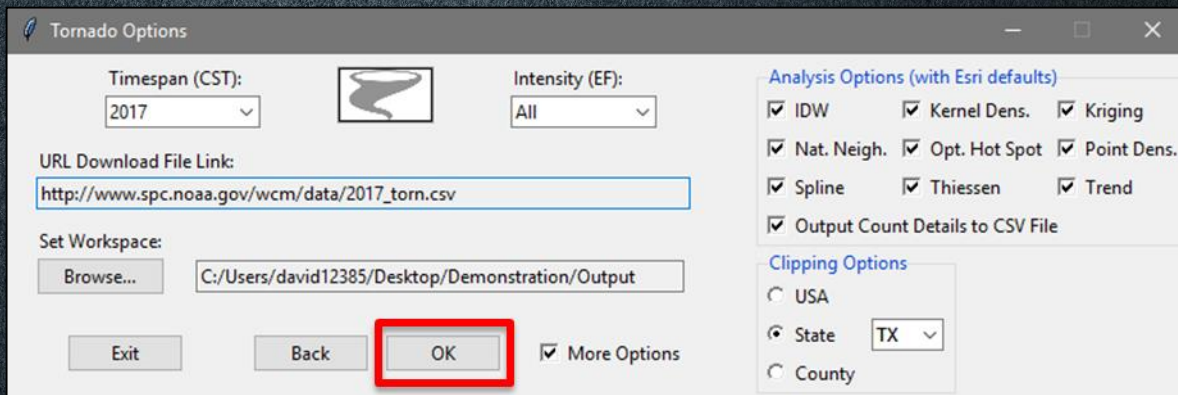


Timespan:	Past 30 Days					
Magnitude Range:	2.5+					
Worldwide Earthquakes:	1716					
USA (and DC) Earthquakes:	287					
State Earthquakes - CA (FIPS: 06):	63					
Statewide Earthquake Data						
V	V	V	V	V		
Minimum Magnitude	Maximum Magnitude	Mean Magnitude	Median Magnitude	Mode Magnitude		
2.45	3.94	2.88	2.82	2.91		
Earthquake Data per County						
V	V	V	V	V	V	V
Counties with Earthquakes	Earthquake Count	Minimum Magnitude	Maximum Magnitude	Mean Magnitude	Median Magnitude	Mode Magnitude
Alpine	1	2.82	2.82	2.82	2.82	2.82
Contra Costa	6	2.48	3.3	2.88	2.93	No Unique Mode
Fresno	4	2.48	2.95	2.73	2.75	No Unique Mode
Humboldt	1	2.58	2.58	2.58	2.58	2.58
Imperial	2	2.48	2.67	2.58	2.58	No Unique Mode
Inyo	1	2.73	2.73	2.73	2.73	2.73
Kern	2	3.69	3.78	3.73	3.73	No Unique Mode
Lake	2	2.91	3.49	3.2	3.2	No Unique Mode
Los Angeles	1	2.5	2.5	2.5	2.5	2.5
Mendocino	2	2.51	2.84	2.67	2.67	No Unique Mode
Mono	3	2.49	3.6	2.89	2.57	No Unique Mode
Monterey	2	2.99	3.11	3.05	3.05	No Unique Mode
Riverside	10	2.45	3.87	2.93	2.91	2.45
San Benito	5	2.6	3.3	2.86	2.82	No Unique Mode
San Bernardino	5	2.84	3.29	3.05	3.01	No Unique Mode
San Diego	5	2.47	3.94	2.88	2.68	No Unique Mode
San Luis Obispo	2	2.61	2.93	2.77	2.77	No Unique Mode
Santa Barbara	1	2.53	2.53	2.53	2.53	2.53
Santa Clara	1	3.82	3.82	3.82	3.82	3.82
Shasta	2	2.52	2.8	2.66	2.66	No Unique Mode
Siskiyou	1	2.8	2.8	2.8	2.8	2.8
Sonoma	4	2.45	2.71	2.61	2.64	No Unique Mode



How Does It All Work?

With all GUI input parameters explained, what happens when the user clicks the OK button?



The screenshot shows the 'Tornado Options' dialog box. The 'Timespan (CST)' is set to '2017', and the 'Intensity (EF)' is set to 'All'. The 'URL Download File Link' is 'http://www.spc.noaa.gov/wcm/data/2017_torn.csv'. The 'Set Workspace' is 'C:/Users/david12385/Desktop/Demonstration/Output'. The 'Analysis Options (with Esri defaults)' section has several checked options: IDW, Kernel Dens., Kriging, Nat. Neigh., Opt. Hot Spot, Point Dens., Spline, Thiessen, Trend, and Output Count Details to CSV File. The 'Clipping Options' section has 'USA' selected, 'State' selected with 'TX' in the dropdown, and 'County' unselected. The 'OK' button is highlighted with a red rectangle.

Tornado Options

Timespan (CST): 2017

Intensity (EF): All

URL Download File Link: http://www.spc.noaa.gov/wcm/data/2017_torn.csv

Set Workspace: C:/Users/david12385/Desktop/Demonstration/Output

Exit Back OK More Options

Analysis Options (with Esri defaults)

- ☒ IDW
- ☒ Kernel Dens.
- ☒ Kriging
- ☒ Nat. Neigh.
- ☒ Opt. Hot Spot
- ☒ Point Dens.
- ☒ Spline
- ☒ Thiessen
- ☒ Trend
- ☒ Output Count Details to CSV File

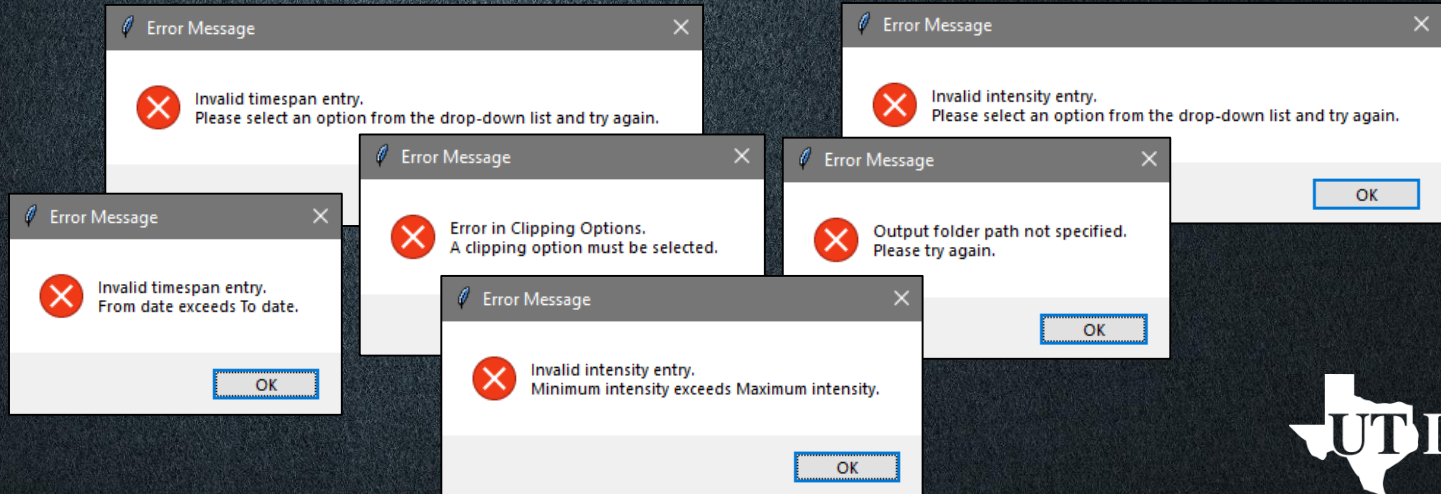
Clipping Options

- ☐ USA
- ☒ State TX
- ☐ County

Parameter Checks

When the **OK** button is clicked, the application checks to ensure all input parameters are valid. Invalid or missing inputs will alert the user to correct any issues.

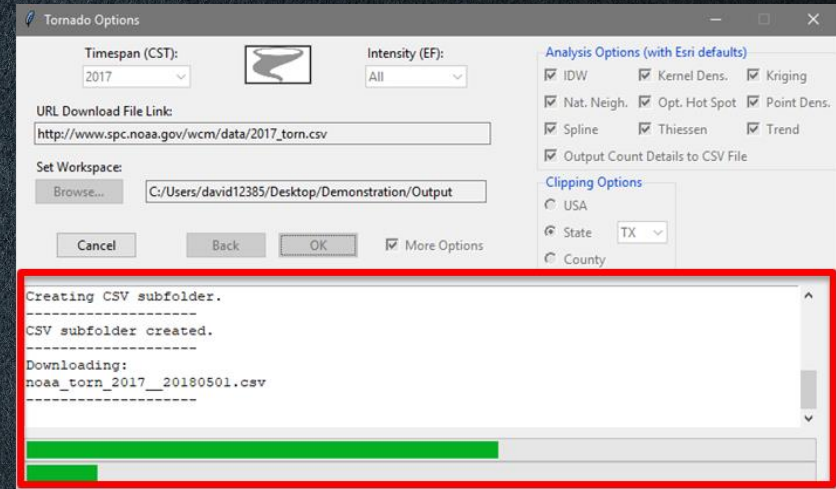
Examples:



If No Errors are Caught...

The GUI will modify itself to show:

- Scroll box with processing messages.
 - Red text = Error
 - Orange text = Warning
 - Blue text = Geoprocessing Message
 - Black Text = Progress messages
- Status Bar
 - Continually loops until finished.
 - Indicates to user it has not frozen.
- Progress Bar
 - Increments to 100%
- All user-selectable items will become disabled.
 - Except Exit button
 - Changes to Cancel



Summarized Processes

1) Creates a parent folder in user-defined output workspace.

- Unique naming conventions per hazard type:
 - {Agency Name}_{Hazard Type}_{Severity}_{Timespan}__{Date of Retrieval}_{Optional Clipping Type}

Name

noaa_torn_all_2017_20180501_state_only_TX

2) CSV/GIS subfolders are created.

3) Raw data is downloaded from user-defined URLs.

Summarized Processes

4) Downloaded raw data goes through quality control checks.

- Rows with missing/invalid Latitude/Longitude are removed.
- Rows with missing/invalid “severity” values are removed.
 - E.g. Negative hail diameters, no tornado EF scale, etc.
- Earthquake rows with man-made, non-natural seismic activity are removed.
 - E.g. Explosions, quarry blasts, etc.
- CSV files with missing headers will have headers added.
- Any unwanted timespans or severity values are removed.
 - E.g. If user only wants EF-3 tornadoes and above, any EF values less than 3 are removed.
- Checked data will have “_checked” added to new CSV file name.
- More specifics within report.



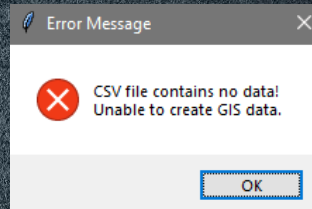
noaa_torn_2017_20180501.csv



noaa_torn_all_2017_20180501_checked.csv

Summarized Processes

- 5) After quality control checks are completed, the remaining data is checked to ensure hazard points/polylines remain.
- If checked data is empty, notify user and terminate process.



- 6) If data exists, convert the CSV/Shapefile into feature class.
- File GDBs created for this, GCS/PCS assigned.

If “More Options” checkbox not selected,
the tool is finished. Most scripts complete in under 2 minutes.

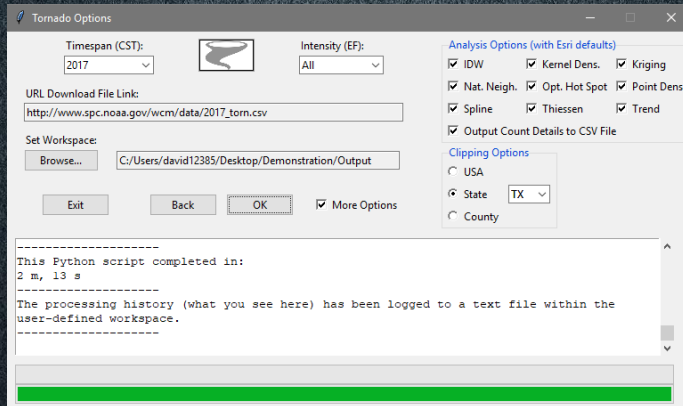
More Summarized Processes

- 7) If “More Options” checkbox selected, the 2017 U.S. Census Bureau Shapefile is downloaded.
 - User-defined clipping extent is then extracted from it.
- 8) The hazard feature class is then clipped to this extent.
 - Saved with “clipped_” in naming convention.
 - If no features exist within clipped output, the user is notified and tool terminates.
- 9) Any selected “Analysis Options” will now be conducted.

More Summarized Processes

10) Once all analyses are complete, the application:

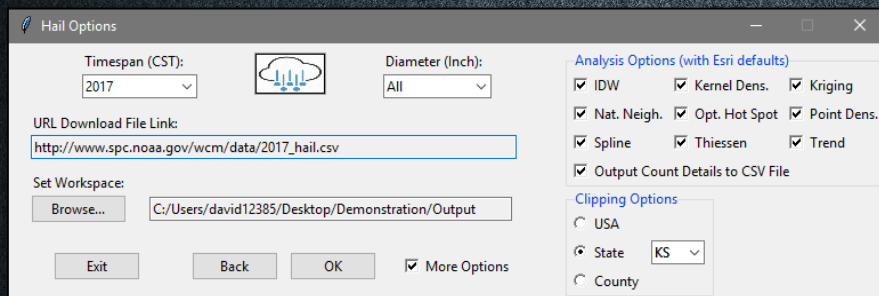
- Shows the elapsed time in scroll box.
- Saves all scroll box messages to text file in parent folder.
- Stops and clears the status bar.
- Sets progress bar to 100%.
- Re-enables all user-selectable items.



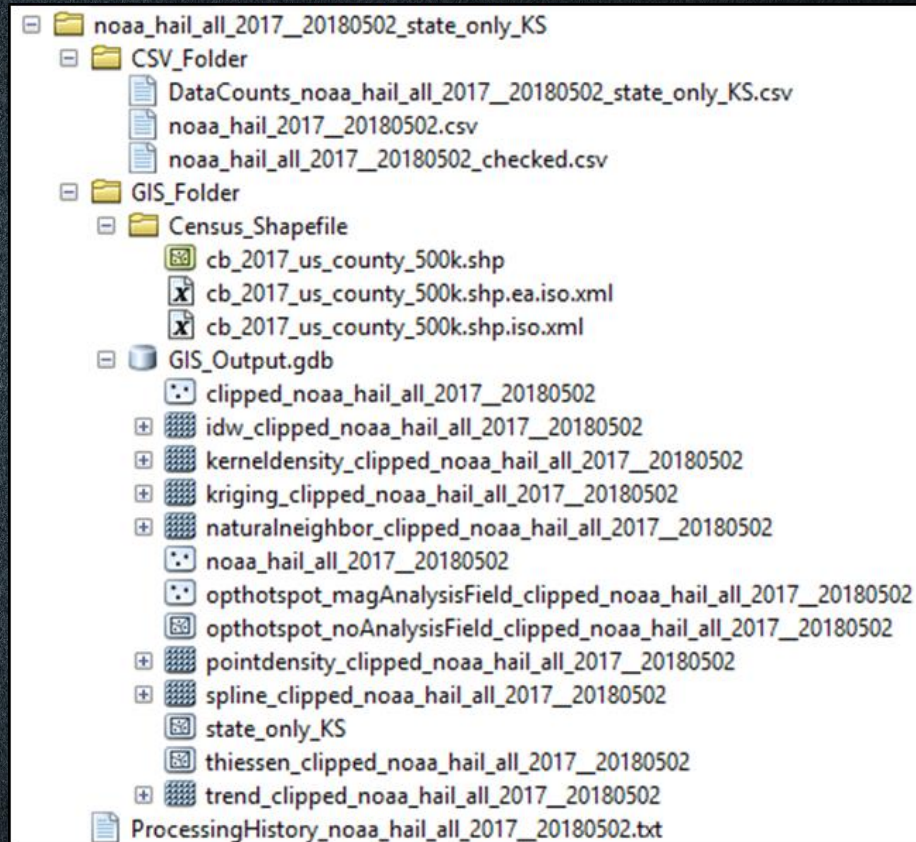
Processing time varies from under 2 minutes to 3 hours, depending on size of dataset selected.

Example Output

All 2017 Hail data for Kansas:

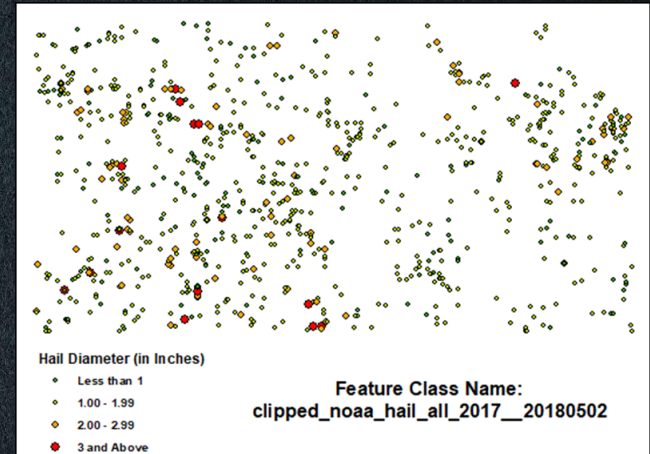
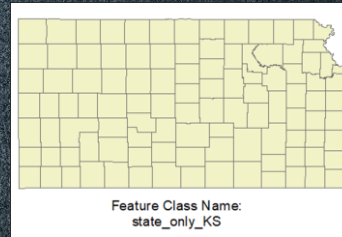
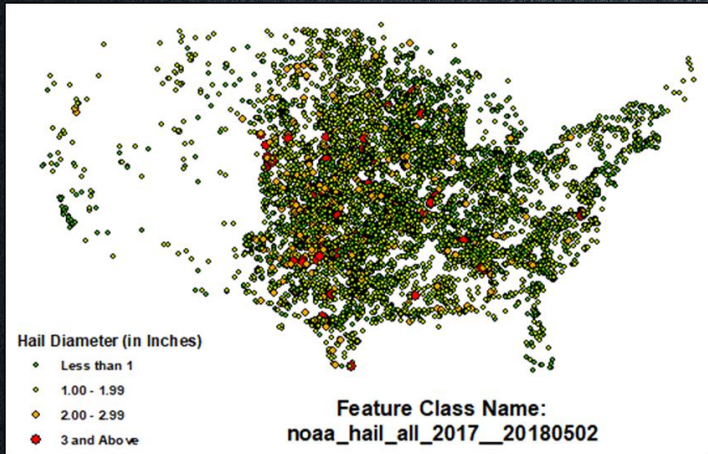


The screenshot shows the 'Hail Options' dialog box. It has several sections: 'Timespan (CST):' with a dropdown set to '2017' and a cloud icon; 'Diameter (Inch):' with a dropdown set to 'All'; 'URL Download File Link:' with a text box containing 'http://www.spc.noaa.gov/wcm/data/2017_hail.csv'; 'Set Workspace:' with a 'Browse...' button and a text box showing 'C:/Users/david12385/Desktop/Demonstration/Output'; and 'Analysis Options (with Esri defaults)' with checkboxes for IDW, Kernel Dens., Kriging, Nat. Neigh., Opt. Hot Spot, Point Dens., Spline, Thiessen, Trend, and 'Output Count Details to CSV File'. There is also a 'Clipping Options' section with radio buttons for 'USA', 'State' (selected, with a dropdown set to 'KS'), and 'County'. At the bottom are 'Exit', 'Back', 'OK', and 'More Options' buttons.



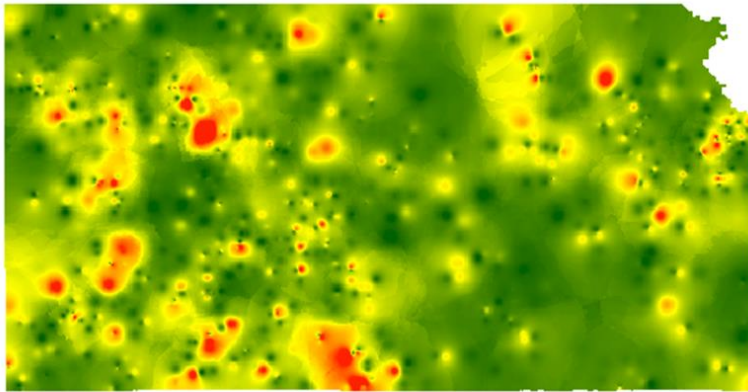
Analysis Output Examples

All 2017 Hail data for Kansas:



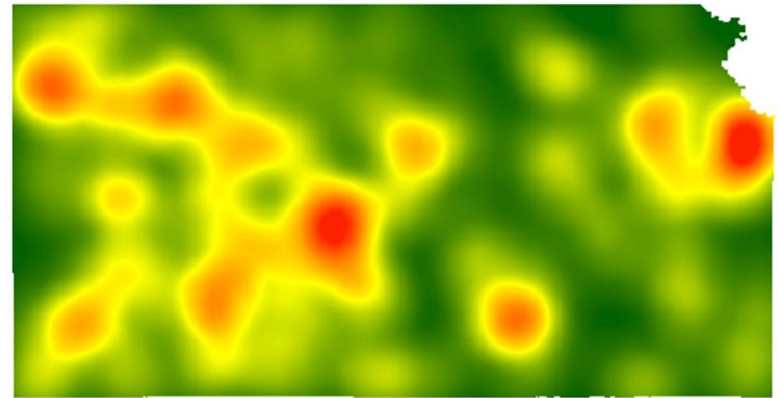
Analysis Output Examples

All 2017 Hail data for Kansas:



Pixel Value
High : 4.24
Low : 0.75

Feature Class Name:
idw_clipped_noaa_hail_all_2017__20180502

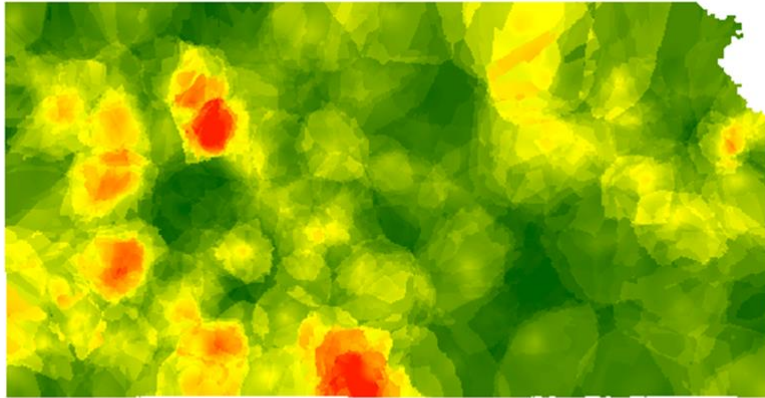


Pixel Value
High : 0.01
Low : 0

Feature Class Name:
kerneldensity_clipped_noaa_hail_all_2017__20180502

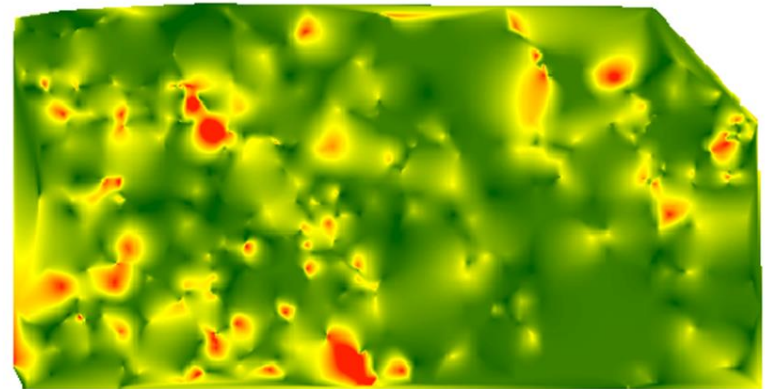
Analysis Output Examples

All 2017 Hail data for Kansas:



Pixel Value
High : 2.52
Low : 0.88

Feature Class Name:
kriging_clipped_noaa_hail_all_2017__20180502

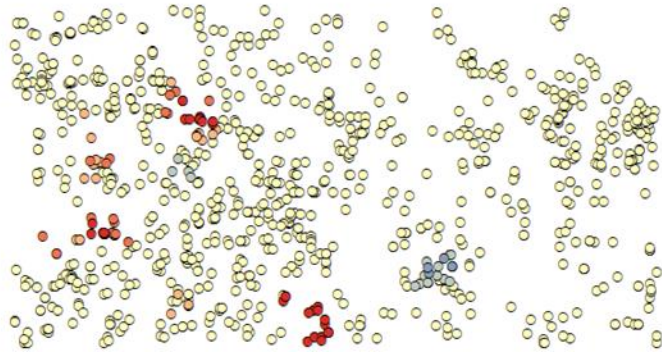


Pixel Value
High : 4.18
Low : 0.75

Feature Class Name:
naturalneighbor_clipped_noaa_hail_all_2017__20180502

Analysis Output Examples

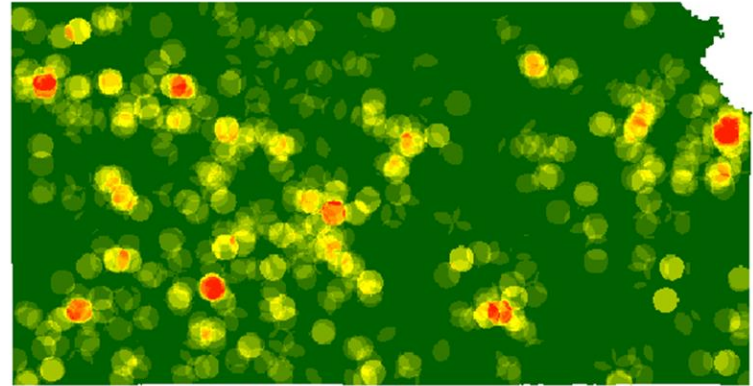
All 2017 Hail data for Kansas:



Feature Class Name:
opthotspot_magAnalysisField_clipped_noaa_hail_all_2017__20180502

Gi_Bin

- Cold Spot - 99% Confidence
- Cold Spot - 95% Confidence
- Cold Spot - 90% Confidence
- Not Significant
- Hot Spot - 90% Confidence
- Hot Spot - 95% Confidence
- Hot Spot - 99% Confidence

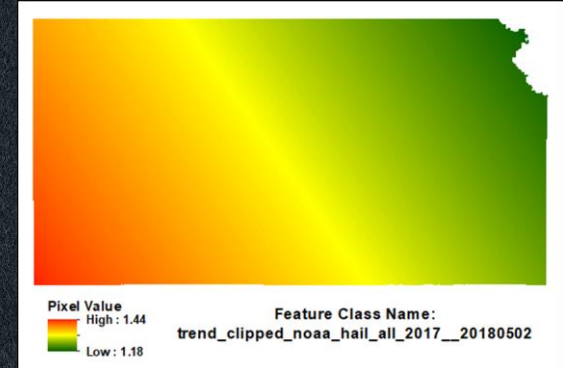
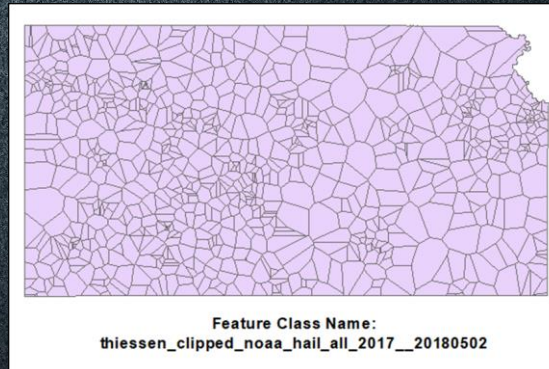


Pixel Value
High : 0.03
Low : 0

Feature Class Name:
pointdensity_clipped_noaa_hail_all_2017__20180502

Analysis Output Examples

All 2017 Hail data for Kansas:



Unique Situation - Earthquake

If user attempts to retrieve earthquake data from custom timespan in excess of 30 days:

- USGS sets a hard limit on data retrieval at 20,000 records.

Error 400: Bad Request

33908 matching events exceeds search limit of 20000. Modify the search to match fewer events.

Usage details are available from <https://earthquake.usgs.gov/fdsnws/event/1>

Request:

/fdsnws/event/1/query.csv?starttime=2018-01-18%2000:00:00&endtime=2018-04-25%2023:59:59&minmagnitude=0&orderby=time

Request Submitted:

2018-04-25T12:26:16+00:00

Service version:

1.5.8

Unique Situation - Earthquakes

20,000 records = ~6-8 weeks' worth of data.

- What if you need several months, or several years??
- To get around this issue, the application is programmed to download one month's worth of data at a time.
- Then proceeds to the next month.
- All monthly data files are then appended into one CSV file.
- Example: Earthquakes from 01/2017 to 03/2017
 - Script downloads 01/2017 data to monthly CSV file...
 - Then 02/2017 data...
 - Then 03/2017 data...
 - The three CSV files are then appended.
 - Application proceeds normally from here.

Unique Situation - Hurricanes

Hurricane data consists of polylines.

- A single polyline path does not accurately represent geographic impact.
- NOAA estimates the hurricane's eye diameter between 20-40 miles across
 - <https://www.hSDL.org/?view&did=34038>
- The eye-wall possesses the most intense wind speeds.
 - With this knowledge, the decision was made to create 50-mile buffer radius for hurricanes to account for size.
 - Clipping extents were also given 50-mile buffer radius.

Final Thoughts

- **Python is more complex than I had assumed.**
 - Application grew to more than 34,000 lines of code!
 - Many challenges faced and were overcome.
- **There is no such thing as a perfect application.**
 - Otherwise we would never receive software updates on our devices.
- **There is always room for enhancements.**
 - Additional hazard types exist and could be implemented to make this application more robust.

End of Presentation

For those interested in having this application, I will provide a Dropbox link for download in the coming days.

Questions?