Workflow to Analyze Hurricane Evacuation Accessibility in Relation to Social Vulnerability

Brendan Sigale | GEOG: 5540

Github Repository

May 10, 2021

Introduction **2**

Data **2**

Original Sources **2**

Data Collection and Descriptive Statistics **3**

Specific Objectives **4**

Script **4**

Inputs 4

Script Pseudocode: 5

Script Details 5

Creating a New Geodatabase 5

Checking Input Feature Data Types 5

Reprojecting Any Necessary Inputs 6

Feature Class Clipping 6

Polygon Centroids 6

Intersect Calculation 7

Near Operation and Routing Discussion 7

Compute Summary Statistics 7

Joining Feature Classes 7

Exporting to a Shapefile 8

Additional Details 8

Discussion **8**

Future Opportunities 8

Conclusion **9**

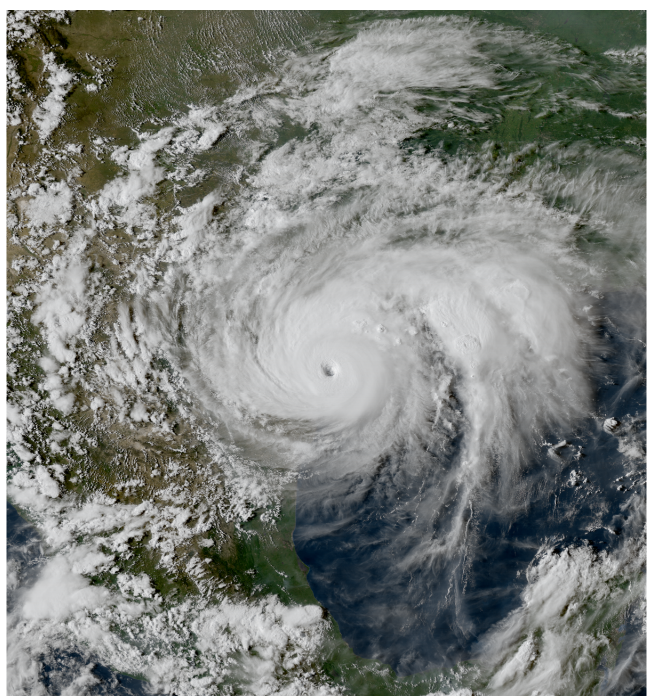
References **10**

Introduction

The purpose of this project is to provide a workflow for the cleaning and processing of hurricane data, social vulnerability, and evacuation route accessibility. Hurricanes can cause immense damage and loss of life, caused by the sustained winds and massive flooding. Texas has previously had issues with evacuation procedures, such as during Hurricane Rita in 2005. According to NPR, Hurricane Rita left 3.7 million Houstonians stranded on the highways in 2005. Because of this ineffective evacuation plan, millions ran out of gas on the highway as the traffic jam stretched for over 100 miles and lasted over 36 hours. Cars running out of gas only made the traffic jam worse, creating gridlock that further slowed the evacuation. Because of this history, Houston Mayor Sylvester Turner and emergency management officials said Houstonians should ride out Hurricane Harvey, as the city is unable to be evacuated. This policy can disproportionately affect those who live in socially vulnerable communities who might be devastated in a hurricane, such as those who live in relatively fragile mobile homes. In such an event, how easy would it be for someone in a socially vulnerable community to evacuate?

To evaluate social vulnerability, this analysis was performed using the CDC’s Social Vulnerability Index (SVI), which ranks census block groups and counties on various social and demographic attributes which can have an affect on how vulnerable the community is, particularly in the event of a natural disaster. This was designed in partnership with the ATSDR's Geospatial Research, Analysis & Services Program (GRASP) to help first responders identify where communities were hit hardest. By integrating additional data, such as historical high water marks and distance to evacuation routes, this response can be augmented to further identify areas which need help most immediately.

Data



Original Sources

This analysis was performed on data from Hurricane Harvey, one of the most devastating hurricanes in recent memory. The map and parallel coordinates plot on the webpage helps understand if there is a relationship between access to a hurricane route and SVI ranking, with some additional information to help interpret what those SVI rankings translate to.

The data is based on a clipping of the SVI from Hydroshare.org containing the SVI information for areas affected by Hurricane Harvey. Hurricane Harvey occurred in August 2017 and caused billions in damage to the Texas, Louisiana, and Alambama coasts, as well as significant damage in the Bahamas. This analysis was performed on data from Texas, as Texas has been notorious in past hurricanes to have millions stranded in highway traffic jams during hurricane evacuations. While this is not a problem exclusive to Texas by any means, Texas is a great location for this analysis as Houston, a very economically and socially diverse large city, was right in the middle of Hurricane Harvey. Understanding disparities in evacuation route access could help urban planners ensure the safety of citizens in their communities.

Also incorporated into this data are US Geological Survey (USGS) recorded high water marks heights, recorded in the immediate aftermath of the hurricane. This data was initially formatted as individual points, but was aggregated by census block group for this analysis. This analysis was performed using ArcPy, which allows the user to group points by outer bounding boxes (in this case, the census block group), and perform summary statistics. For this analysis, the average height of high water marks in each census block group was used.

Additional data includes TxDOT's Open Data Portal road inventory shapefile, formatted as a polyline, which contains all roads in the state. Also used is TxDOT's evacuation route shapefile, which contains designated emergency evacuation routes from the coast further inland. It contains detailed information about what type of evacuation route each road is designated as.

Data Collection and Descriptive Statistics

- CDC Social Vulnerability Index
    - Shapefile polygons containing SVI's data.
    - Contains Census data and CDC recorded metrics, including ranking, for all census tracts impacted by Hurricane Harvey. This includes tracts in Texas, Louisiana, Mississippi, Arkansas, and Alabama, totaling to 6,439 rows and 130 columns.
    - The top 10% are ranked as 1 and the bottom 10% are ranked as 0 to increase vulnerability of these tracts.
    - Tracts which do not have population data are ranked as -999, and will be removed from the data in the workflow.
    - Each census tract's percentile is stored in the field 'RPL_THEMES'
    - https://www.hydroshare.org/resource/c2df2a80b9d6490788704a24854f4879/

- Texas Evacuation Maps
    - Shapefile Polylines containing the route name, route type, shape length, and global ID.
    - Contains 164 Polylines, made up of Major Evacuation Roads, Potential EvacuLanes, and Potential Contraflow. EvacuLanes are areas which, when activated, allow the use of the road's shoulder. Contraflow, when activated, reverses a road's direction, transforming inbound lanes into outbound lanes.
    - The data contains 139 Major Evacuation Routes, 13 Potential EvacuLanes, and 12 Protential Contraflows.
    - https://gis-txdot.opendata.arcgis.com/datasets/txdot-evacuation-routes?geometry=-117.399%2C26.556%2C-75.586%2C33.218

- Texas Roadways
  - Shapefile Polylines containing a road's identifying features, location, type, direction, and current status.
  - Contains 222,215 polylines.
  - https://gis-txdot.opendata.arcgis.com/datasets/d4f7206d27af4358acb70cb1cc819d10_0?geometry=-96.579%2C28.970%2C-95.926%2C29.075

- USGS Hurricane Harvey High Water Marks
  - Contains point data of recorded water peaks along the Texas Coast.
  - Contains 1,258 points and 43 fields.
  - Relevant field for analysis is peak_stage, which is the recorded water height.
  - The average peak_stage is 85.44, and the distribution skews heavily towards minimal values. That being said, there are still hundreds of values which are well above the mean, with the max being 453.8 feet.
  - https://www.hydroshare.org/resource/615d426f70cc4346875c725b4b8fdc59/

Specific Objectives

Some of the specific objectives this workflow will be able to answer (and will be answered in this analysis:

- Do census tracts which rank low on the SVI have better access to designated evacuation routes?
- Do census tracts which rank higher on the SVI flood more easily?

These specific objectives were chosen because they can be assessed programmatically using the readily accessible data from Hurricane Harvey. In addition, they can be effective indicators of correlations between social vulnerability and how a community will be able to handle emergencies. By assessing how far a census block group is from an emergency evacuation route and how likely that area is to flood, it can be used to determine if an area needs better access to evacuation routes or repairs.

Script

The data was cleaned using a Python script, developed in ArcGIS Pro. The script performs multiple tasks all at once to ensure consistent and smooth operation and output. It will output each polygon input with the desired columns*, summary statistics, and euclidean distance to the nearest evacuation route access point. The output is a shapefile in the defined path with the defined output name.

Inputs

- svi - Path to polygons, in this analysis representing census block groups.
- evac - Path to polylines, in this case representing TxDOT's emergency evacuation routes.
- roads - Path to roads, in this case using TxDOT's road inventory database.

- hwm - Path to points, in this case representing USGS high water marks from Hurricane Harvey. Can be any points.
- workspace - Path to folder where new analysis gdb should be created.
- statField - Field which summary statistics will be crreated from. Must be in hwm.
- statType - Summary statistic type. Must be 'SUM', 'MEAN', 'MIN', 'MAX', or 'STDDEV'.
- joinField - Unique field in svi which will be used to join the analyses back together.
- output_path - Path of desired output shapefile.
- output_name - Name out output shapefile.
- fieldmap - see Additional Details

These inputs allow for a detailed and consistent analysis using multiple input features. All inputs, except for fieldmap, are required for the function to perform appropriately.

Script Pseudocode:

1. Create a new database in a defined workspace.
2. Check input feature data types.
3. Reproject any unmatching feature classes to that of the roads input.
4. Clip data into the necessary boundaries.
    a. This is first done by clipping the hwm input to svi, and then doing the reverse using a spatial join on svi.
5. Centroids are calculated for the input polygons to estimate distance to roads.
6. Intersect analysis is performed between roads and evac inputs to estimate evacuation route access points.
7. Distance between polygon centroids and the nearest intersection is calculated.
8. Specified summary statistics are computed for the provided inputs.
9. Data is joined together using the input polygon's unique identifier.
10. Data is exported to a feature class, then to a shapefile, using the user-defined output information and field mappings.

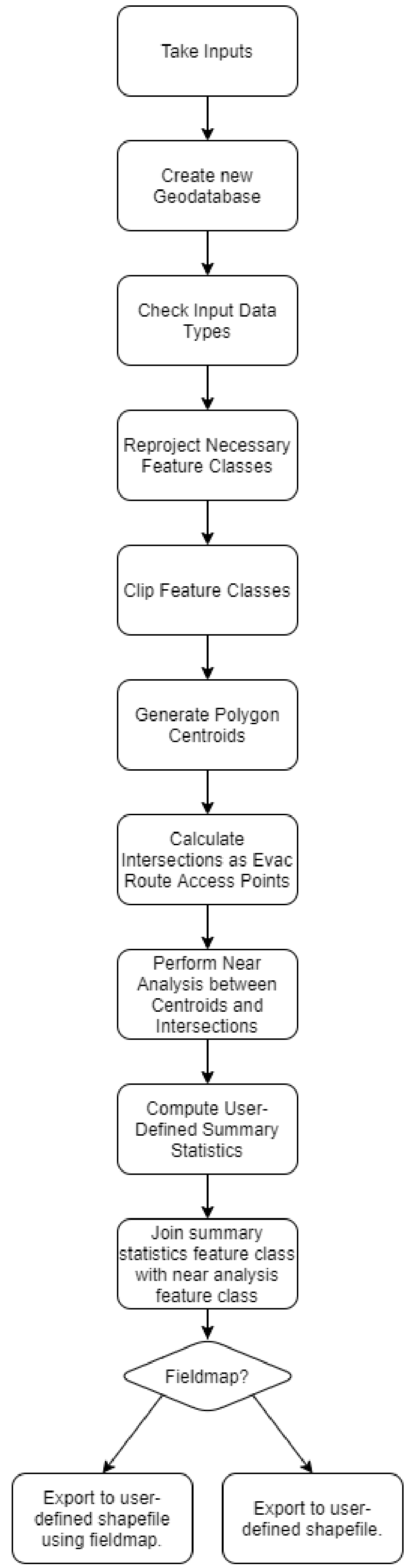
Script Details

Creating a New Geodatabase

This section of the code creates a new geodatabase to contain all analysis layers. It uses the defined user workspace. Note: the workspace path cannot be a geodatabase.

Checking Input Feature Data Types

This section of the code checks all input feature class types to ensure they meet the appropriate standards. This is accomplished by describing each input object and then checking the shapeType. If they do not match the designated input type, a TypeError is raised indicating which input has issues and what shapeType it should be.



Reprojecting Any Necessary Inputs

This section checks to ensure all input types match and reprojects any necessary inputs to match the roads feature class. It accomplishes this by using the describe objects created in an earlier section and checks to make sure the projection names match the roads feature class. The roads feature class was chosen arbitrarily. If they do not match, an indicator message is printed that they do not match and what projection the unmatching feature class will be updated to. The reprojection is then performed using arcpy.management.Project(). The file path originally passed as an input is then updated to point to the reprojected feature class to ensure it can be easily referenced in later sections of code.

Feature Class Clipping

Feature classes are then clipped to the study area. This is done in two steps to limit all features to only the study area. First, the hwm input points are clipped to only the census block study areas.

Next, a spatial join was performed between the input polygons and the clipped hwm point feature class. This limits the polygon feature class to only the study area where points were located, both to limit the output data size and to ensure no errors were encountered when calculating summary statistics later in the script.

These steps were necessary in this data because there were both points and polygons outside of the study area. Polygons extended deep into Texas where there were no recorded high water marks, and there were high water marks outside of Texas, which was not part of the study this analysis was designed for. A user who would like to not clip out points should remove these sections and be careful to update the object references to ensure ‘hwm_clipped’ and ‘svi_clipped’ point at the original input files.

Polygon Centroids

This block calculates centroids for each census tract to generalize the near analysis later. This is performed using arcpy.management.FeatureToPoint(), with the third parameter set to “CENTROID”.

Intersect Calculation

Intersection points are calculated between roads and evacuation routes to identify evacuation route access points. This is performed using arcpy.analysis.Intersect() and outputs a point feature class.

Near Operation and Routing Discussion

Euclidean distance is calculated between the croid of each polygon and the nearest evacuation route access point using arcpy.analysis.Near(), with the third parameter being "GEODESIC". This returns the euclidean distance to the nearest evacuation route on ramp.

Ideally, this section of the code would have returned the amount of minutes it would take to reach the access point the day before the hurricane made landfall at Noon. This would use a Find Closest Facility Analysis Layer using the road input feature class. However, this part of the analysis did not work, as the only output was in meters. In addition, the majority of the input polygon feature class centroids did not have a route generated for them. The reason for this is unknown, but the end result was incomplete data, so this was not possible. The Near tool's output was the best remaining option for this analysis, as it returned outputs for each input feature.

Compute Summary Statistics

This section of the code computes summary statistics on the input point feature class within each census block group using arcpy.Analysis.SummarizeWithin(). The user is able to pass in a field they would like to generate statistics for and what statistic they would like. They have the option between sum, mean, minimum, maximum, and standard deviation.

Before computation, the inputs are checked. The statField input parameter is checked against the field names in the point feature class using list comprehension and arcpy.ListFields(). If the input field name is not in the list of fields, a NameError is raised. The statType input parameter is also checked against a list of possible statistic options. If the input statType is not in this list, a NameError is raised.

There is currently no specific check for if the input field is a numeric. Instead, the script attempts to compute the summary statistics. At this point in the code, everything works except for if the field is non-numeric. If the SummarizeWithin function fails, an exception is raised that indicates the likely issue that that statField is non-numeric.

Joining Feature Classes

The data is joined back together to account for the separation between the polygon feature class used in the summary statistics analysis and the centroids generated for the near analysis. This takes the input joinField to join the data together. This only needs to be a unique identifier in the svi input feature class. The same field is used for both because only the shape type changes - all data between the two feature classes is the same. This is performed using arcpy.analysis.AddJoin().

Exporting to a Shapefile

The data is converted to a feature class, which then is output as a shapefile in the designated output path with the designated name. Column outputs can be specific using the fieldmap input parameter, but this may require two runs and access to Arc tools to be effective. See more in additional details.

Additional Details

Exporting specific columns requires the use of field mappings. As described in the script documentation, acquiring field mappings requires this analysis to essentially be run twice. The first time it is run will be without a field mappings parameter. This will export the same analysis, but include all columns. The user should then use Arc's Feature Class to Feature Class tool to create a new feature class with all desired columns. Once this is completed, the user should then copy the field mapping from the results tab into a string, which can then be passed into the function to be run again. This will then output a shapefile with the user's desired columns. While this workaround is strange, it is necessary as several joins occur, creating hundreds of fields with different names and unique identifiers. Once this is done once, the data can be changed to any desired spatial area or points, as long as the columns the user wants to keep have the same names. This allows for some form of repeatability, and is better than the alternative of an output shapefile with hundreds of columns, many of which duplicate one another due to the multiple joins.

Discussion

Future Opportunities

In the future, integrating a more detailed analysis of evacuation accessibility, such as a drive-time analysis, would be ideal in the future. However, this was not possible due to limitations in the scripting and other unknown limitations which did not allow the data to be processed.

If these errors were to be resolved, it would be interesting to see accessibility differences between different areas and how long it took them to get out of the storm's predicted path the day before it made landfall. This would provide greater insight into possible inequalities, such as whether or not a road has adequate capacity to handle the population quickly evacuating. For example, a census block group which is low risk in the SVI may have wider and higher quality roads than one which is high risk. Thus, they may have as far of a distance to access an evacuation route, but take significantly less time because the roads can handle more throughput than roads in a socially vulnerable census block group.

An example of a census block group which successfully generated a route (blue square with brown line), next to ones which did not (brown squares).

An example of a census block group (blue square) which did not find any routes, despite being along networked roads and near evacuation route access points.

The image to the above shows how certain census block group centroids were simply unable to find acceptable routes, despite being right next to them. The roads were all networked properly using the Arcpy's network analyst package, but routes could not be generated, in some cases even for census block groups right next to one another.

In this analysis specifically, the USGS High Water Marks sometimes contained sporadic measurements, especially in the downtown Houston area, so some census block groups are not represented. In the future, it would be better to have these additional values to help further understand the relationships between poverty, flood damage, and access to evacuation routes. There is another shapefile containing points of buildings predicted by FEMA to be damaged and modeled inundation and damage. However, these points covered less tracts than the High Water Marks and integration between the two datasets was not successful. Ideally, a weighted average would be calculated between the mean water height and mean modeled inundation in each census block group based on the amount of points each dataset had within its boundaries. However, calculating this analysis was not successful.

Conclusion

This script is effective but could use further tuning in the future. For example, it could be made to be more general than the existing workflow to allow for more functionality, such as multiple summary statistics. In addition, it could integrate with traffic data to further identify inequalities in hurricane evacuation. Regardless, this workflow and function delivers analysis which could be useful for further analysis. It has already been useful for visualization in the associated Geographic Visualization project. This workflow lays the groundwork for future analysis which can be used by urban planners and emergency management personnel to ensure safety for their citizens. In addition, it can allow first responders during or after a hurricane to identify which communities may have gotten stuck during evacuation. This is one of the additional goals of the Social Vulnerability Index, so adding additional data to the analysis only further helps this cause.

References

(USGS), U. G., & Arctur, D. (2018, November 26). USGS - Harvey High Water Marks. Retrieved from https://www.hydroshare.org/resource/615d426f70cc4346875c725b4b8fdc59/

Domonoske, C. (2017, August 28). Why Didn't Officials Order The Evacuation Of Houston? Retrieved from http://www.npr.org/sections/thetwo-way/2017/08/28/546721363/why-didn-t-officials-order-the-evacuation-of-houston

TxDOT Evacuation Routes. (n.d.). Retrieved from https://gis-txdot.opendata.arcgis.com/datasets/txdot-evacuation-routes?geometry=-124.431,26.556,-68.554,33.218

TxDOT Roadways. (n.d.). Retrieved from https://gis-txdot.opendata.arcgis.com/datasets/d4f7206d27af4358acb70cb1cc819d10_0?geometry=-99.471,28.568,-92.486,29.409

U.S. Centers for Disease Control and Prevention / Agency for Toxic Substances and Disease Registry / Geospatial Research, A. (2018, November 26). CDC Social Vulnerability Index 2014. Retrieved from https://www.hydroshare.org/resource/c2df2a80b9d6490788704a24854f4879/