SSCI Project 1

Sandeep Kahlon

University of Southern California

Prof. Swift

Introduction

Project one aimed to enhance students' foundation in spatial and proximity analysis through the application of Arcgis Pro and the integration of Python scripting. The project's overarching goal was to classify private and public schools in LA within a 1-mile radius of fault lines as Class A-Class D on basis of relative earthquake risk. The slip rate of fault lines, defined as the rate at one side of the fault slides past the other side, was leveraged in order to assess and classify nearby schools' earthquake risk. Higher slip rates tend to have more frequent earthquakes and historically earthquakes resulting from faults with lower slip rates are scarce (CA Conversation 2022). Earthquake risk is a predominant natural phenomenon that occurs in the region of Southern California, through appropriate classification of schools on a uniform basis with respect to fault lines can benefit preventative and safety measures. This report outlines several geoprocessing tasks, scripted in python to manipulate and transform the geospatial shapefile datasets retrieved from the California Department of Conservation in order to obtain the desired and accurate classification of schools relative to earthquake risk.

Study Area

The study area concerns Los Angeles county located in the southern region of California. Los Angeles county is approximately 4,753 square miles in area and contains a population of 9.83 million. As California is regarded as the American state with the highest occurrence of earthquakes per capita and the high density of earthquake faults in the southern region, LA county serves satisfactory for the conjunct school and earthquake spatial analysis.

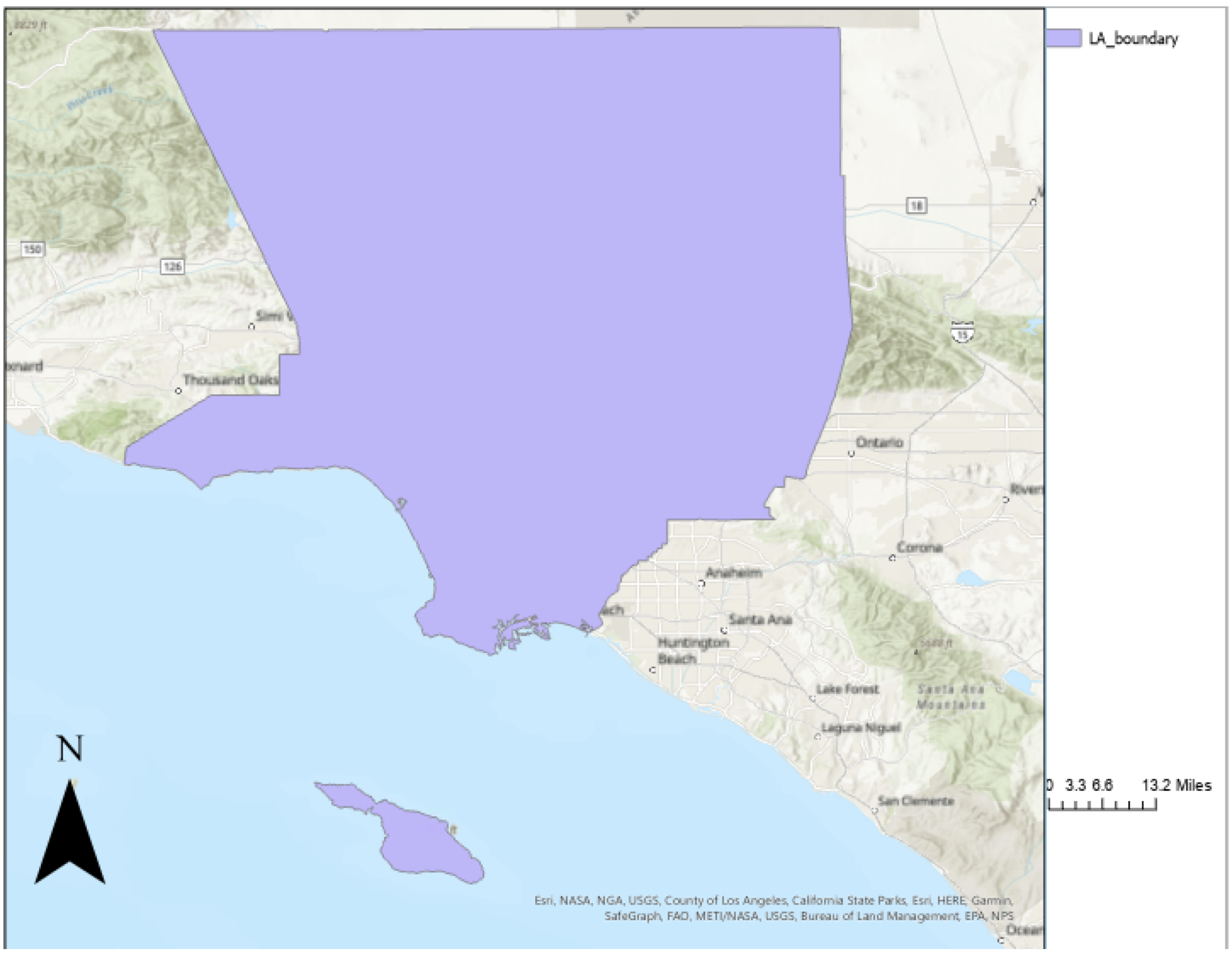

Figure 1 - The study and spatially analyzed region of Los Angeles County, California.

Data and Data Processing

Dataset	Status	Description
LA_boundary	Incorporated	***Data Type:*** Shapefile. ***Characteristics:*** Polygon features (1 row, 1 column). ***Geographic Range:*** Los Angeles County. ***Spatial Reference:*** WGS 1984 Web Mercator (PCS). ***Comments:*** Details the geographic extent of the County of Los Angeles, California.
EarthquakeFaults_USGS_CA_v2	Incorporated	***Data Type:*** Shapefile. ***Characteristics:*** Line Features (4,224 rows 33 columns) ***Geographic Range:*** The entirety of the State of California, United States. ***Spatial Reference:*** WGS 1984 Web Mercator (PCS). ***Comments:*** Pinpoints the location and extent of fault lines in California along with fault line characteristics inclusive of name, slip rate, etc. Retrieved from USGS.
EarthquakeFaults_USGS_LosAngeles	Not Used	***Data Type:*** Layer File. ***Characteristics:*** Line Features (4,224 rows 33 columns). ***Geographic Range:*** The entirety of the State of California, United States. ***Spatial Reference:*** WGS 1984 Web Mercator (PCS). ***Comments:*** Retrieved from the USGS, this dataset builds upon the

Dataset	Status	Description
		EarthquakeFaults_USGS_CA_v2 dataset as it classifies fault lines (Class A-C) with respect to earthquake occurrence.
Public_Schools	Incorporated	***Data Type:*** Shapefile. ***Characteristics:*** Point Features (10,551 rows, 21 columns). ***Geographic Range:*** The entirety of the State of California, United States. ***Spatial Reference:*** WGS 1984 (GCS). ***Comments:*** Provides geographic locations and descriptive features of public schools in the state of California.
Private_Schools	Incorporated	***Data Type:*** Shapefile. ***Characteristics:*** Point Features (1,514 rows, 18 columns). ***Geographic Range:*** The entirety of the State of California, United States. ***Spatial Reference:*** WGS 1984 (GCS). ***Comments:*** Provides geographic locations and descriptive features of private schools in the state of California.

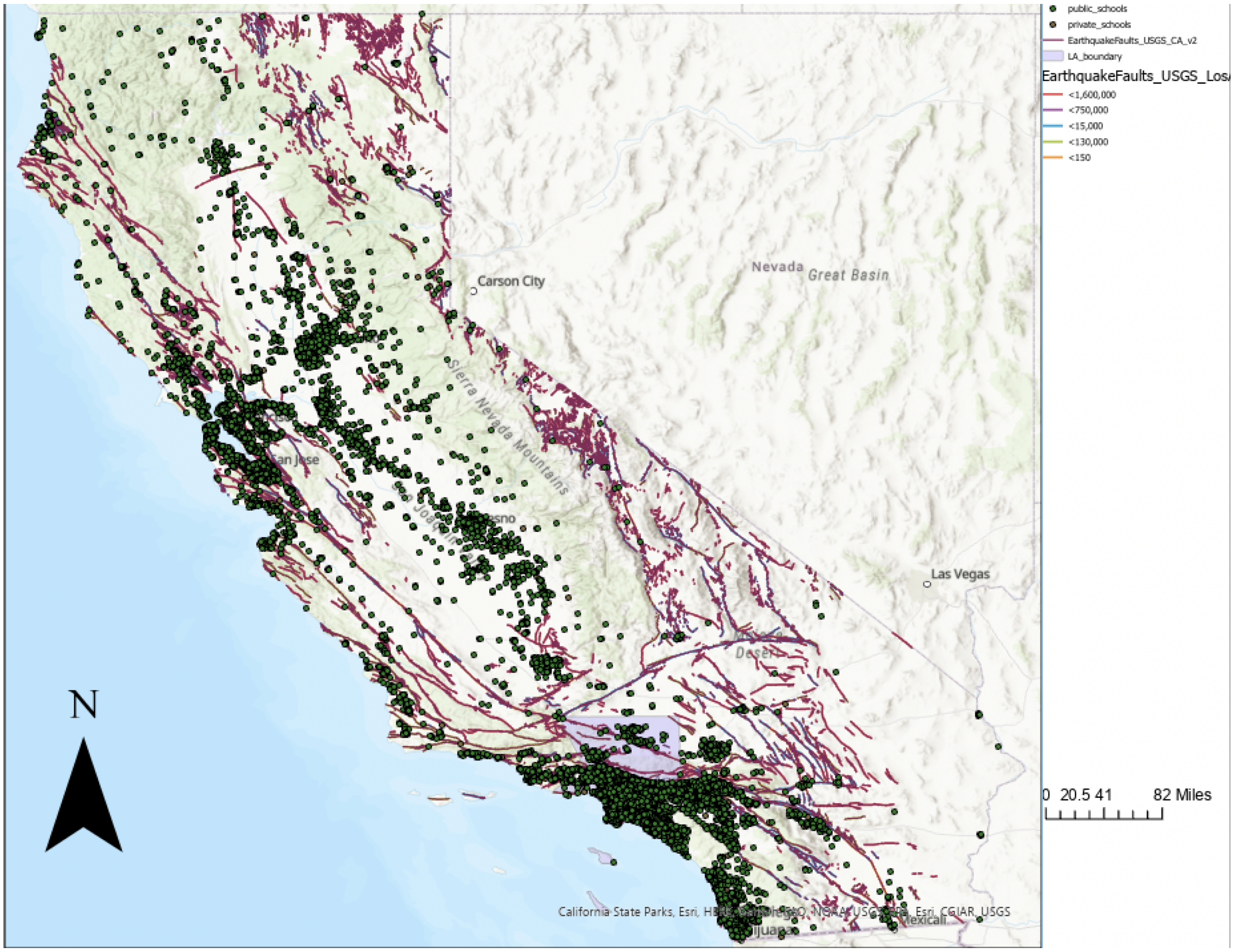

Figure 2 - Original data layers (4 shapefile, 1 layer file) projected in the map in original respective projected and geographic coordinate systems.

Methodology

To execute the respective spatial and proximity analysis mentioned, the application of ArcGIS and Pycharm editor was required. A new project labeled 'Project 1' was initiated in ArcGIS and the five shapefile and layer file datasets provided respecting schools, earthquake faultlines, and the Los Angeles County boundary were imported into the project's geodatabase. Following, the spatial datasets were all converted to a homogeneous projected coordinate system (WGS 1984 Web Mercator) by incorporating the ArcGIS project tool. Thereafter, Python scripting language was utilized through the application of Pycharm editor to perform the spatial analysis and geoprocessing tasks required for the project.

```
import arcpy
arcpy.env.workspace = 'G:\proj1_586\proj1_586.gdb'

#Question 1

fc_ls = ['private_schools', 'public_schools', 'EarthquakeFaults_USGS_CA_v2']
for val in fc_ls:
    arcpy.Clip_analysis(val, 'LA_boundary', f'{val}_LA')

#Question 2

arcpy.management.Merge(['public_schools_LA', 'private_schools_LA'], 'schools_LA')

#Check
arcpy.GetCount_management('schools_LA')
#<Result '3062'>
print(int(arcpy.GetCount_management('private_schools_LA')[0]) + int(arcpy.GetCount_management('public_schools_LA')[0]))
#<Result '3062'>

#Question 3
arcpy.Buffer_analysis('EarthquakeFaults_USGS_CA_v2_LA', 'EarthquakeFaults_LA_Buffer', '1 MILE')

arcpy.Clip_analysis('schools_LA', 'EarthquakeFaults_LA_Buffer', 'danger_schools')

#Question 4
#obtain information regarding which faultline is closest to respective schools
arcpy.analysis.Near('danger_schools', 'EarthquakeFaults_USGS_CA_v2_LA', method='Planar')

#Join field to reference closest fault line slip-rate with respect to LA Schools
arcpy.management.JoinField('danger_schools', 'NEAR_FID', 'EarthquakeFaults_USGS_CA_v2_LA', 'OBJECTID_1')

#Add Empty field to append class data
arcpy.management.AddField('danger_schools', 'Risk_Level', 'TEXT')

#Append class data to empty field on basis of slipcode
with arcpy.da.UpdateCursor('danger_schools', ['slipcode', 'Risk_Level']) as cursr:
    for row in cursr:
```

```
    if row[0] == 4:
        row[1] = 'Class D'
        cursr.updateRow(row)
    elif row[0] == 3:
        row[1] = 'Class C'
        cursr.updateRow(row)
    elif row[0] == 2:
        row[1] = 'Class B'
        cursr.updateRow(row)
    elif row[0] == 1:
        row[1] = 'Class A'
        cursr.updateRow(row)
```

Figure 3 - Python code executing the four geoprocessing required tasks for project completion.

With the application of Pycharm editor, a new python script file was created and a workspace connection was made with the ArcGIS project geodatabase that was initiated so geoprocessing tasks could be executed upon the spatial datasets. The arcpy library was imported for utilization of various geoprocessing and data analytic modules. Initially, to constrain all data to the Los Angeles county study area, the feature data layers were clipped to the Los Angeles county boundary data layer through the arcpy Clip_analysis tool as visualized in question one of figure 3. A for loop was initiated to iteratively conduct the geoprocessing task for memory optimizatoin. In order to evaluate schools on a uniform basis with respect to relative earthquake risk, the merge tool was utilized to merge public and private school datasets into a new feature class labeled 'schools_LA' as stated in question 2 of figure 3. As the project required assessing earthquake risk of only schools within 1 mile of fault lines, a buffer around the USGS earthquake faults data layer (EarthquakeFaults_USGS_CA_v2) was developed through utilizing arcpy's Buffer_analysis module with a parameter of "1 mile". Thereafter, the school features were clipped once again to the created fault line buffer zone through applying the Clip_analysis tool. Both of these geoprocessing tasks are scripted under question 3 in figure 3.

With schools being constrained to the sub geographic desired (1-mile from faultlines in Los Angeles county), the schools were classified with respect to earthquake risk. In the fault line data layer a column detailing slip rate exists, I selected this feature to be deterministic of earthquake risk as a high slip rate suggests a high rate of earthquakes. The near_analysis arcpy module was leveraged on the schools and fautline data layers to append a column to the school's data layer detailing the object ID of the nearest faultline. Thereafter, to conduct spatial analysis of faultlines' slip rates and schools in conjunction, these two datasets were joined on the object ID primary key. A new empty column in the school data layer was developed through the usage of the arcpy AddField module. A conditional if-loop was created to append Class A-D values on basis of the slip rate (>2.5 = Class A, .5-2 = Class B…). to the empty column as shown in question 4 of figure 3. Following these geoprocessing and data manipulation steps, schools were correctly classified (Class A-D) on basis of the earthquake risk sourcing from the slip rate of the nearest proximity fault line. Finally, appropriate unique value symbology was placed upon the school features with respect to their earthquake risk classification.

Results

Through executing the python script demonstrated above in Figure 3, the following results in Figures one and two were obtained. Only schools in the Los Angeles county boundary were accounted for and they were provided symbology which details their classification of earthquake risk (Class A-D). As there remained no values of schools near a fault line with a slip rate of <.5 and slip code of 1, there failed to be any schools classified as Class D. Figure 1 visualizes the output of program and the developed buffer zone where schools were clipped. Figure 2 displays the output, it is evident that only schools within one mile of the fault zone were respected in the analysis. Clusters of schools with respective classifications were also observed throughout the study region of Los Angeles County.

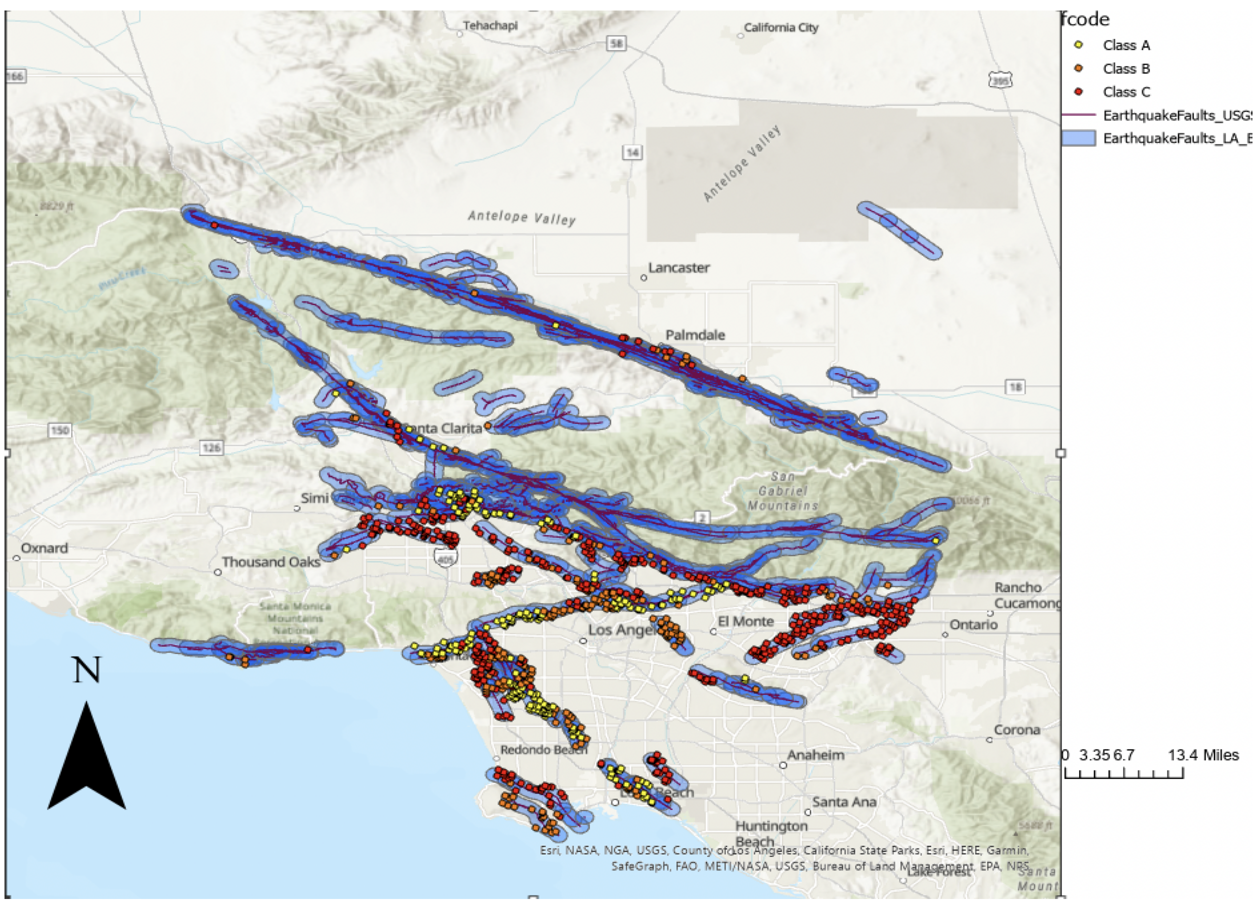

Figure 1 - Visualization of the final output resulting from geoprocessing tasks scripted in python aiming to classify schools in LA county within a 1-mile radius from fault lines with respect to earthquake risk. Visualized with buffer.

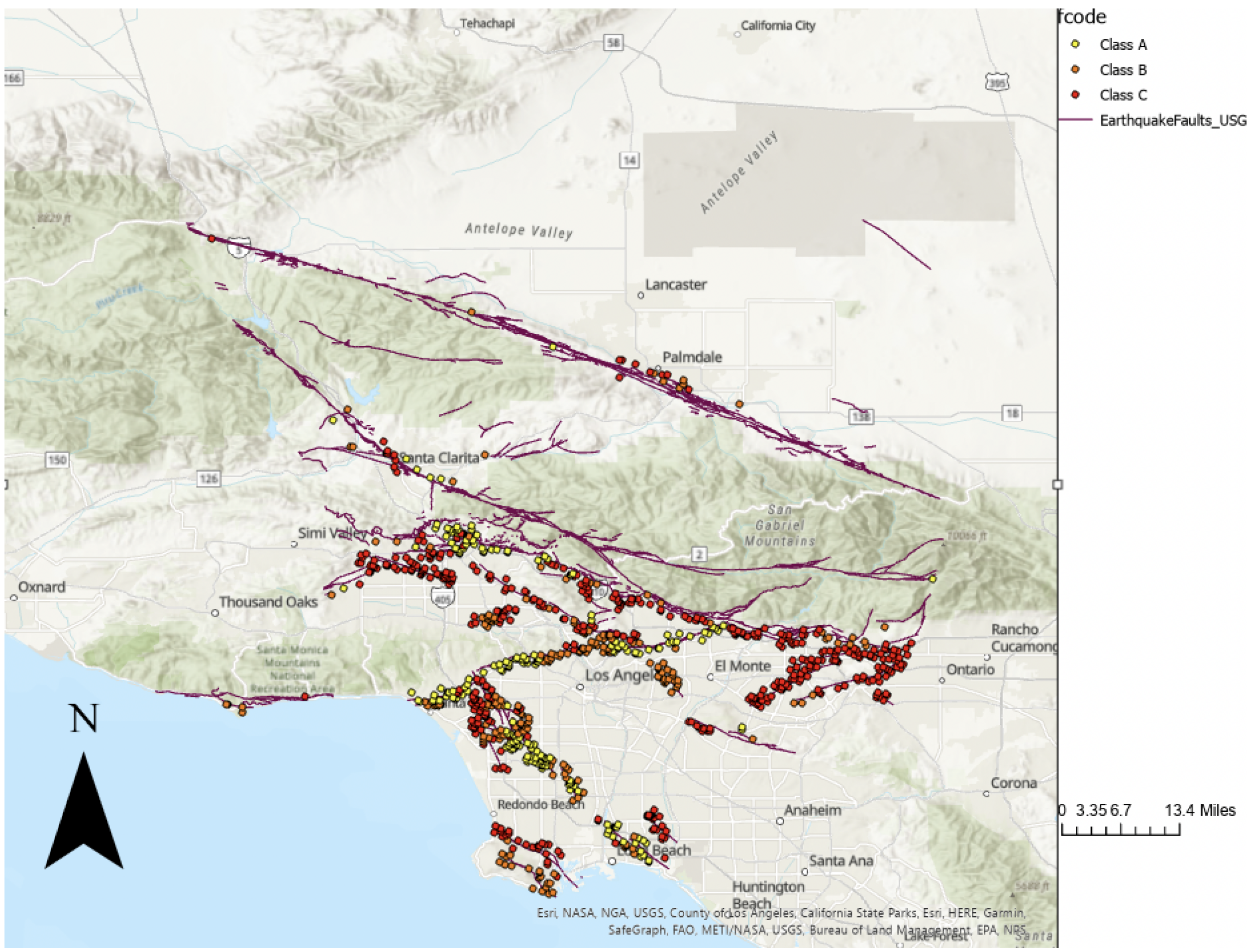

Figure 2 - Map result of the final output resulting from geoprocessing tasks scripted in python aiming to classify schools in LA county within a 1-mile radius from fault lines with respect to earthquake risk. Visualized without buffer.

Discussion

With Los Angeles County recognized as the highest populated county in the state of California alongside the state's historical record of earthquakes, this project classifying schools near a 1-mile distance of fault lines on basis of earthquake risk can emit several solutions. As several clusters of classified schools were evident in the finished map, preventative measures and resources can be allocated more efficiently to regions in Los Angeles county where there exist clusters of schools with high earthquake risk. The results concur with the study conducted by (Husid et.al 2007, 8-9) where the researchers concluded that pre-disposed knowledge and preventative measures regarding earthquakes can reduce the distribution of colossal damage that occurs to schools in the LA region.

Through scripting in Python and leveraging the arcpy library, several geoprocessing tasks were able to be automated through control and conditional loops. This served as beneficial for time capacity and memory storage. ArcGIS served as the platform where output results and data layers project upon the map with the execution of the associated Python code. Importing of data and symbology configuration was also completed in ArcGIS.

Resources

Conservation, California Department of. “Fault Slip Rates.” CA Department of Conservation. Accessed January 22, 2023. https://www.conservation.ca.gov/cgs/Pages/PSHA/sliprates.aspx.

Husid, R., A. F. Espinosa, and J. de las Casas. “The Lima Earthquake of October 3, 1974: Damage Distribution.” Pubs.geoscienceworld.org, 2007. https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/67/5/1441/117774/The-Lima-earthquake-of-October-3-1974-Damage.