Spatial Data Integration and Analysis at the California Institute of Technology

Chenyi Weng

SSCI581- Concepts for Spatial Thinking

Wang

March 19, 2024

1. Introduction

This report outlines the process of acquiring and preparing spatial data for analyzing the geographical and infrastructural aspects of the California Institute of Technology (Caltech) campus. Moreover, utilizing ArcGIS Pro, four distinct datasets from diverse sources were collected and processed, focusing on tree coverage, building structures, traffic collisions, and demographic patterns within the Caltech vicinity. Eventually, the goal was to integrate these datasets to provide a comprehensive spatial analysis framework for future studies.

2. Evaluation of the Sample Project

The sample project, different sets of data related to the area surrounding the University of Southern California's University Park campus were looked at. In addition, building footprints from OpenStreetMap (OSM) and LA GeoHub were found to have different ways of categorizing features, showing that not all maps indicate buildings in the same way. However, when checking addresses that had been marked with their location on a map against these building footprints, it was clear that they didn't always match up perfectly, which shows how tricky it can be to combine different kinds of map data. Moreover, the study of trees and surfaces that water can't go through (such as concrete) over time showed how the area has changed, similar to how knowing when buildings were built can give clues about the area's history. Furthermore, adding pictures to the data made it easier to see and understand how buildings, green spaces, and hard surfaces are spread out across the area. At last, if more information could be added to this study, things like who lives in the area or how the land is used could give a fuller picture of how the area works and changes.

3. Study Areas and Coordinate System

The study focuses on the Caltech campus, nestled within Pasadena, California. However, the defined boundary for this analysis was carefully selected to encompass the campus and immediate surrounding areas. Also, the chosen projected coordinate system, NAD 1983, facilitates accurate local scale analyses. It is worth noting that this system ensures consistency across datasets, enabling precise spatial analysis within the study area.

4. Data Acquisition and Preparation

4.1 The Building Dataset of Caltech

Data from OpenStreetMap (OSM), covering tree coverage as of February 2024, was chosen for its detailed representation of urban greenery.[1] The process involves searching for "Caltech" on OSM, exporting the selected area's data in .osm format, and converting it into a .gdb file for use in ArcGIS as shown in Figure 1. Furthermore, the dataset, initially in the WGS 1984 coordinate system, is projected into the NAD 1983, using the Project tool, ensuring compatibility with other datasets. As shown in Figure 2, this dataset was adjusted to match the campus boundary and reprojected for consistency with the study's coordinate system. Finally, a visual representation of this dataset within ArcGIS Pro highlights the spatial distribution of trees across the campus.

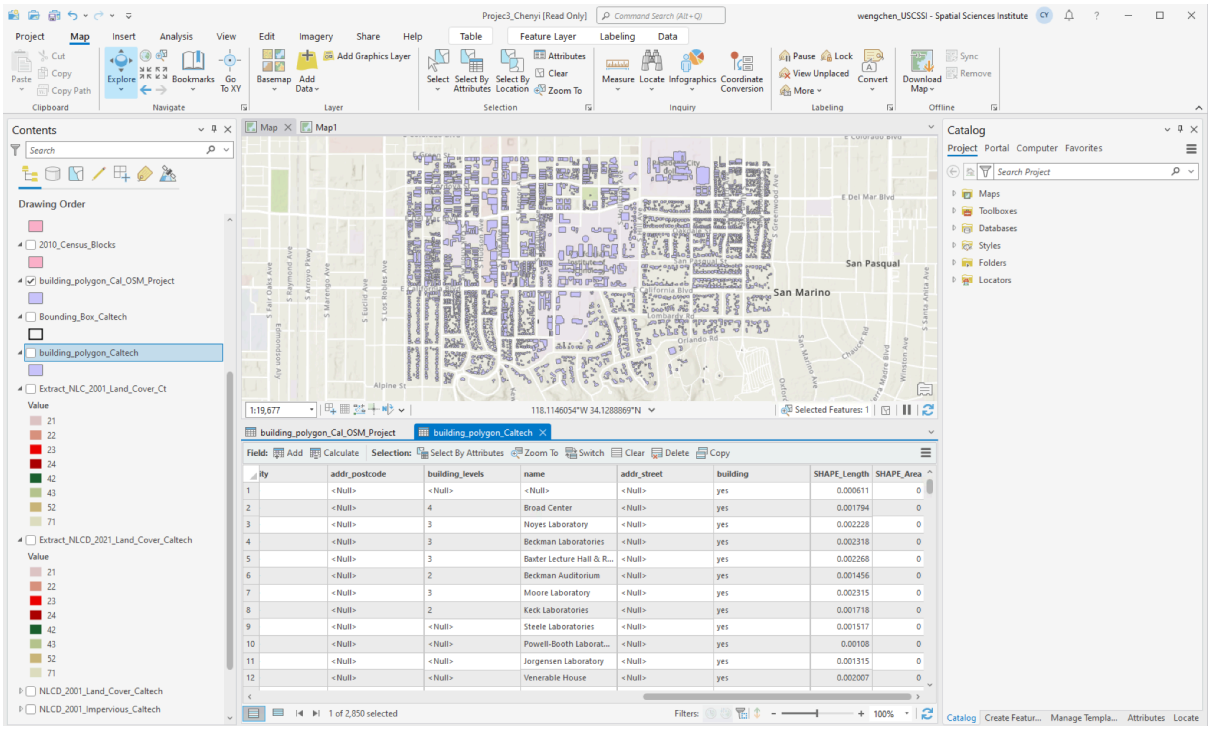


Figure 1. Screenshot of the building polygon dataset before setup

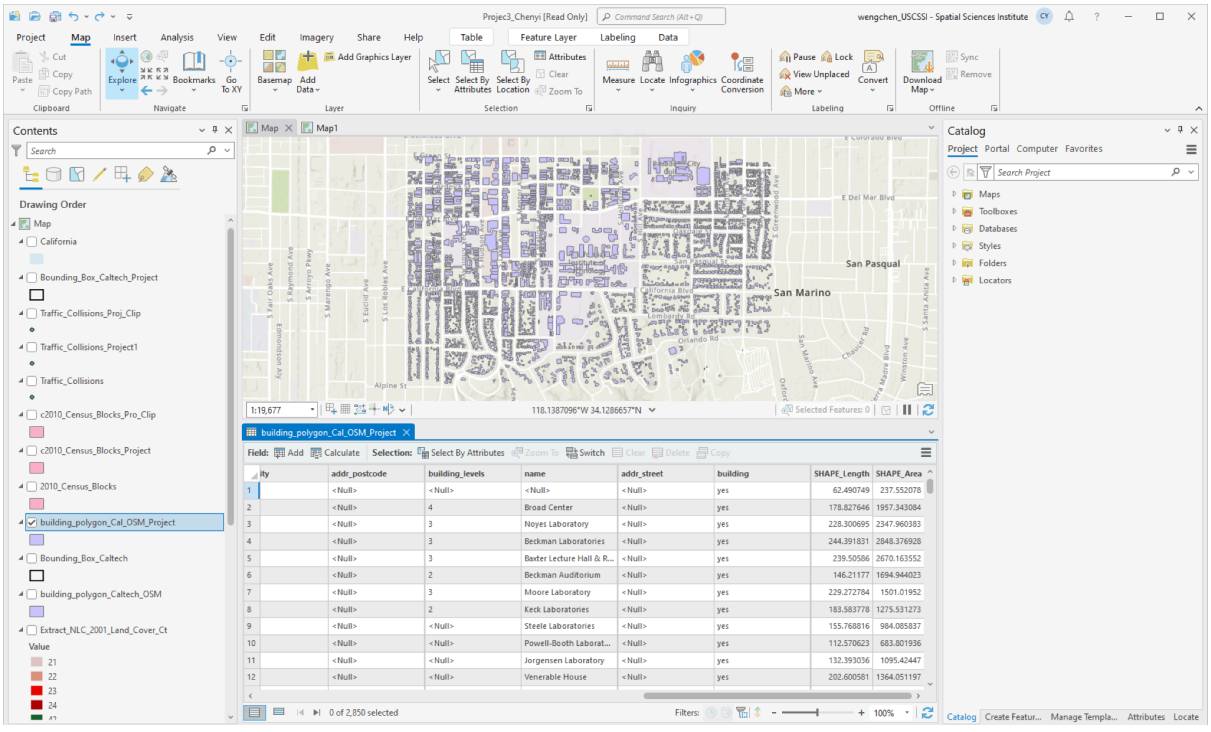


Figure 2. Screenshot of the building polygon dataset after setup

4.2 The Land Cover Dataset of Caltech

As shown in Figure 3 and Figure 4, building structures were mapped using the Caltech Building Polygon dataset acquired from the MRLC Consortium for 2001, and 2021. After clipping to the study area, this dataset was projected into the selected coordinate system, revealing the layout and density of campus buildings; also, as shown in Figure 5 and Figure 6, after importing the raster data into ArcGIS, the Project tool adapts its coordinate system to match the other layers. This means that a bounding box is created around Caltech for clipping the data to the campus area using the Extract by Mask tool. However, land cover data for 2001 and 2021 was obtained from the MRLC Consortium to analyze environmental changes over time.

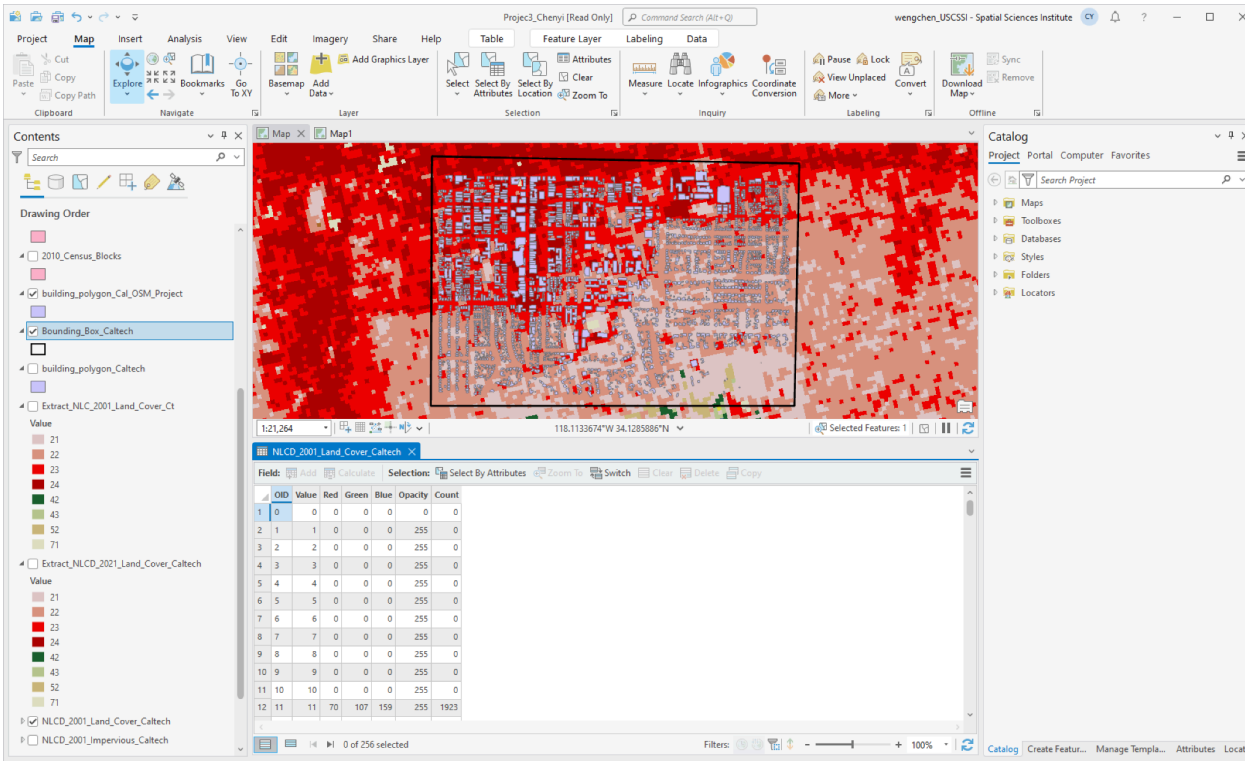


Figure 3. The NLCD Land Cover dataset (2001) before setup

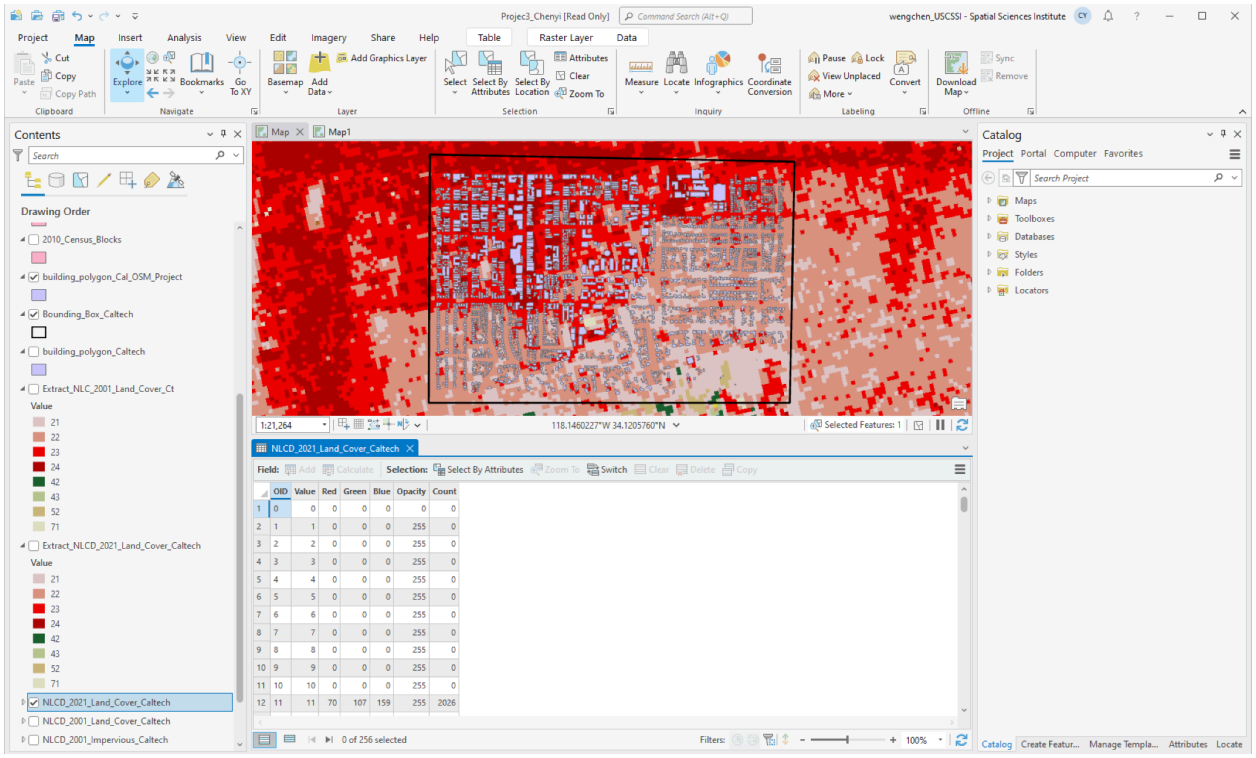


Figure 4. The NLCD Land Cover dataset (2021) before setup

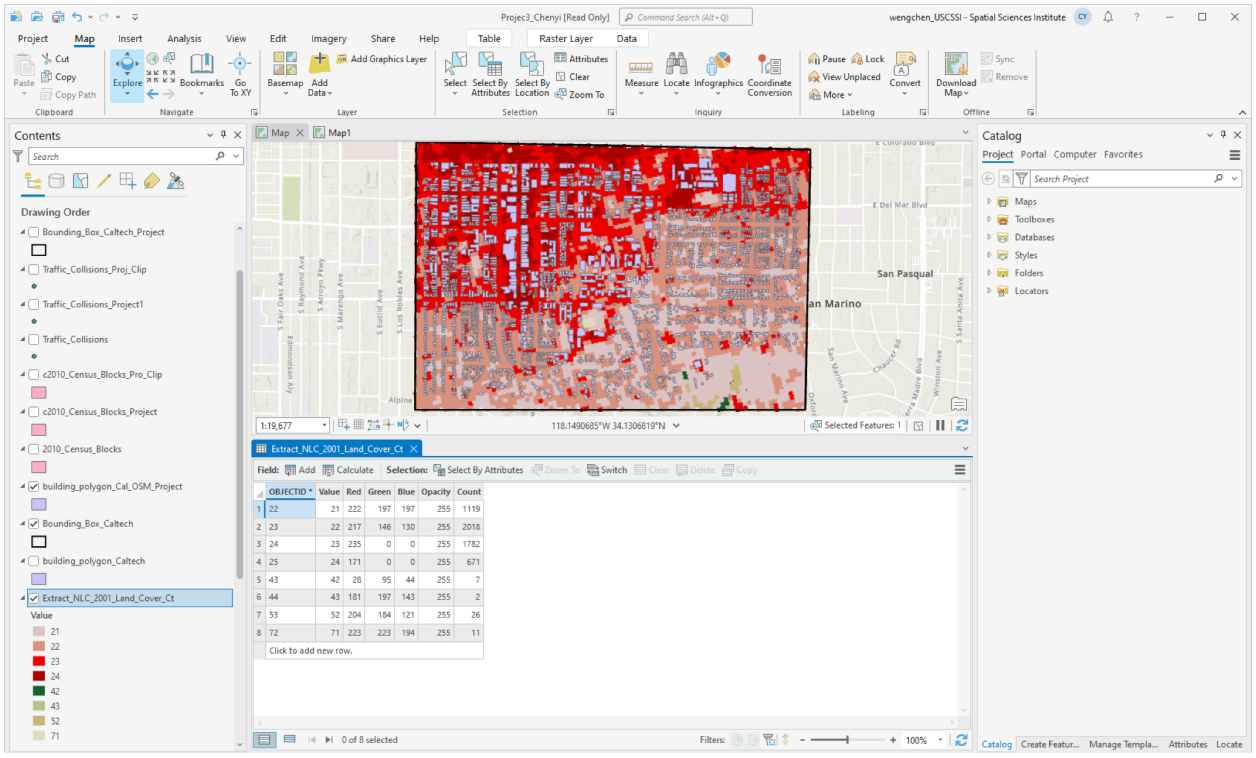


Figure 5. The NLCD Land Cover dataset (2001) after prepared

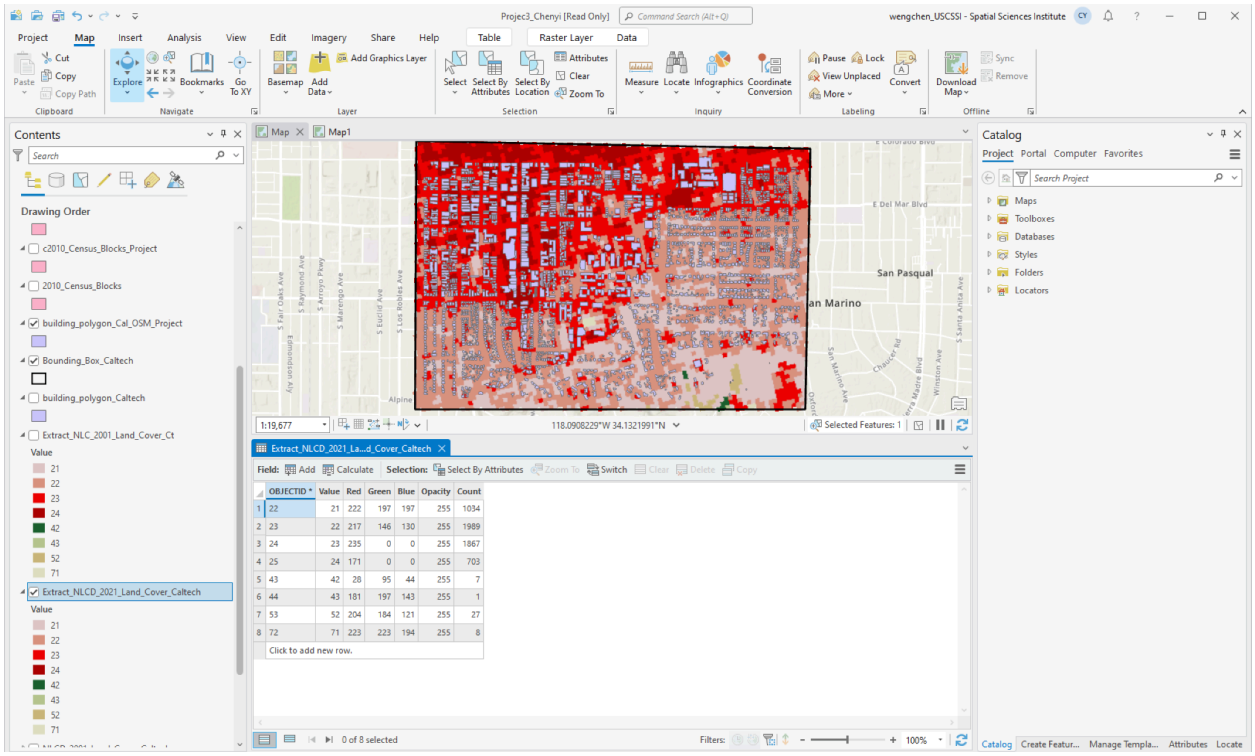


Figure 6. The NLCD Land Cover dataset (2021) after setup

4.3 Traffic Collisions Dataset

As shown in Figure 7, traffic collision data, provided by the City of Pasadena Open Data Site as of May 5, 2022, underwent similar processing. Namely, the shapefile is imported, and clipped to the study area after being projected into the correct coordinate system as shown in Figure 8. Ultimately, this dataset was instrumental in identifying patterns and areas of concern regarding vehicular incidents around the campus.

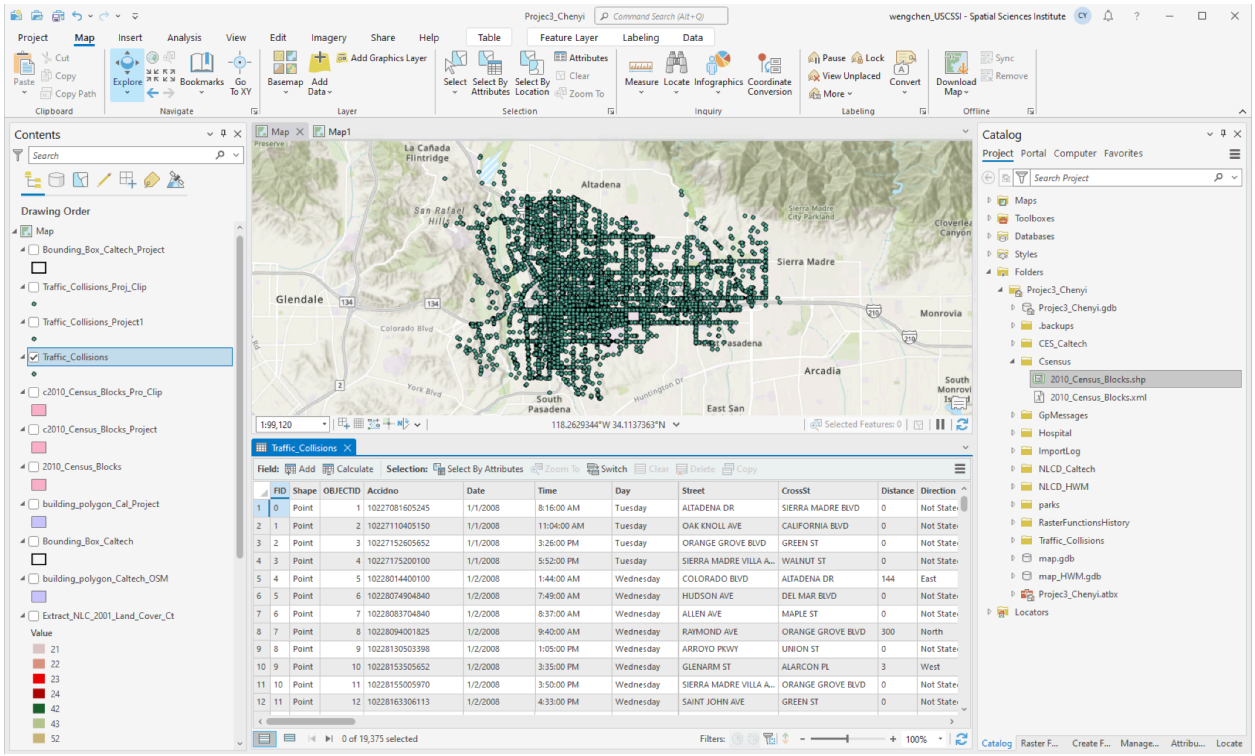


Figure 7. The Traffic Collisions dataset before prepared

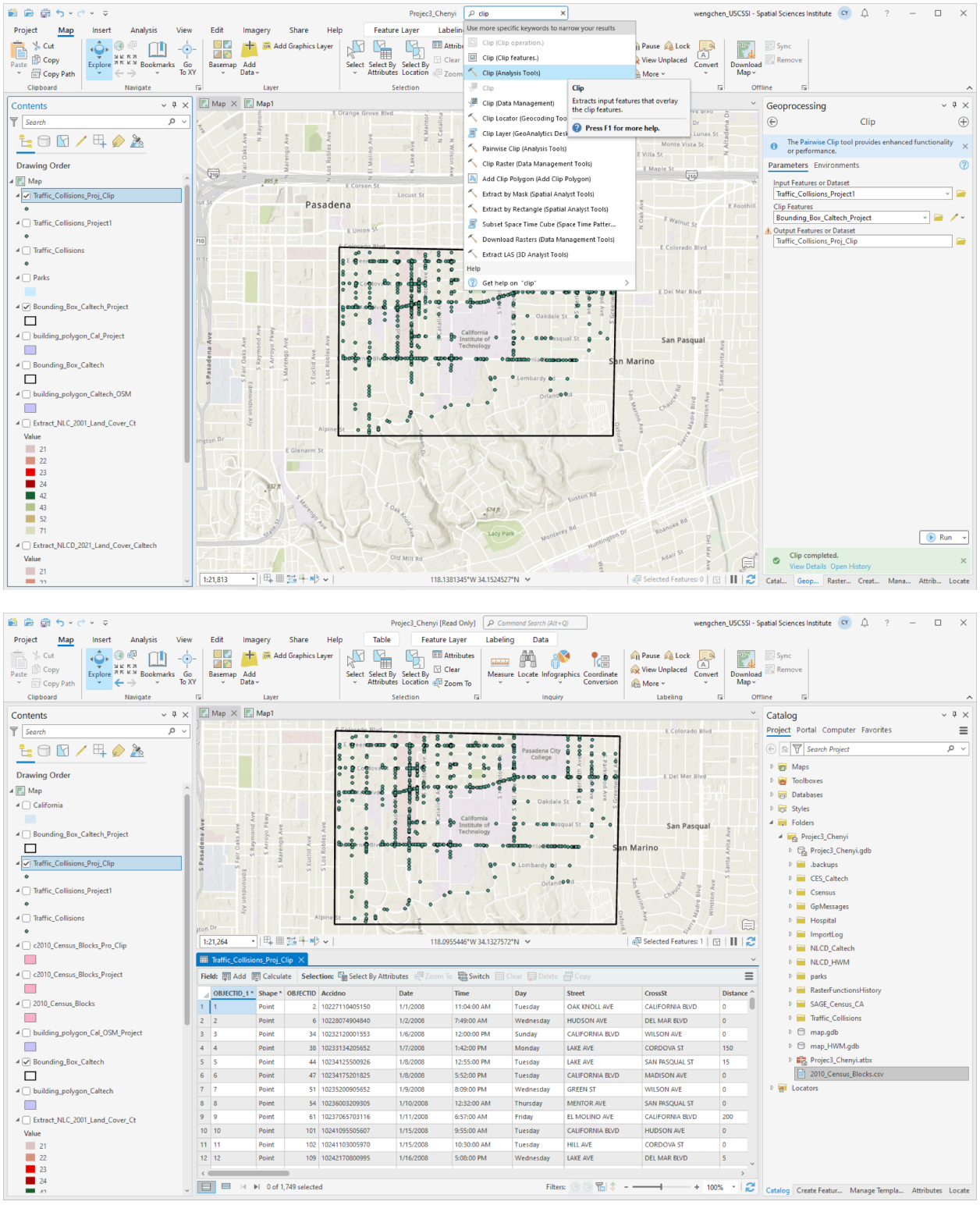
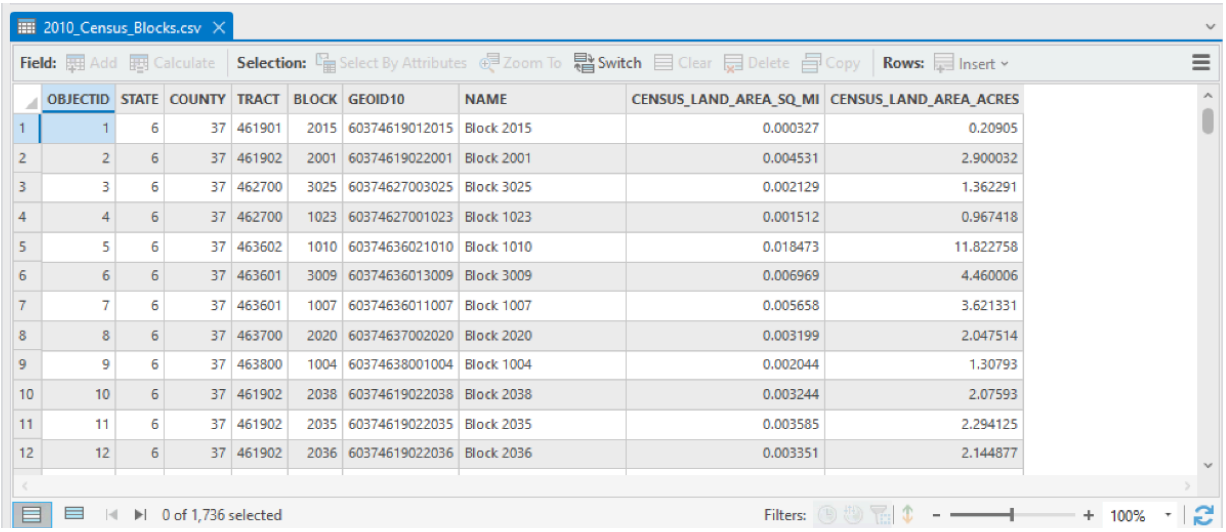


Figure 8. The Traffic Collisions dataset after prepared

4.4 2010 Census Blocks Dataset

Lastly, demographic information from the 2010 Census, sourced from Sage Data, enriched the future study with socio-economic context. First of all, the original .csv data is imported into ArcGIS using the Geocode Table tool as shown in Figure 9, converted into a .gdb format as shown in Figure 10, and then clipped to the study area after projecting into the suitable coordinate system as shown in Figure 11. After a long time, this data was aligned with the campus area and provides insights into the population distribution around Caltech.



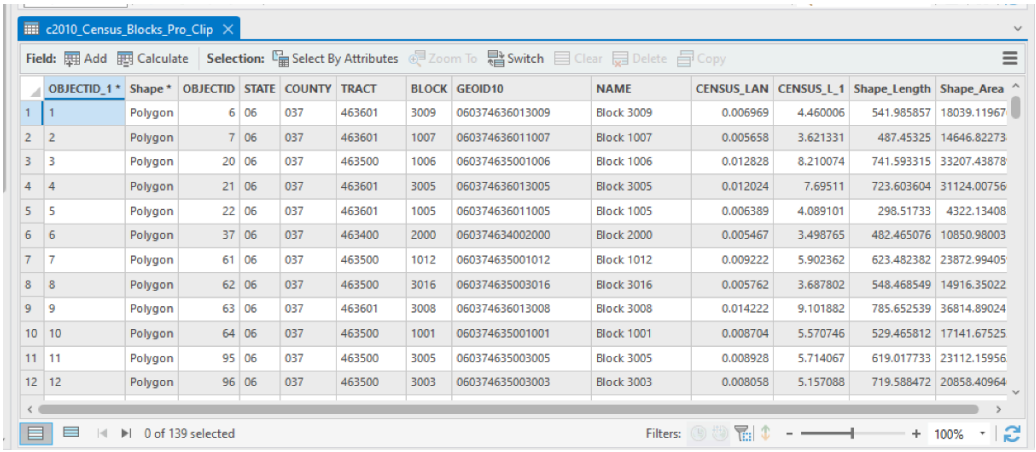
2010_Census_Blocks.csv

Field: Add Calculate | Selection: Select By Attributes Zoom To Switch Clear Delete Copy | Rows: Insert

	OBJECTID	STATE	COUNTY	TRACT	BLOCK	GEOID10	NAME	CENSUS_LAND_AREA_SQ_MI	CENSUS_LAND_AREA_ACRES
1	1	6	37	461901	2015	60374619012015	Block 2015	0.000327	0.20905
2	2	6	37	461902	2001	60374619022001	Block 2001	0.004531	2.900032
3	3	6	37	462700	3025	60374627003025	Block 3025	0.002129	1.362291
4	4	6	37	462700	1023	60374627001023	Block 1023	0.001512	0.967418
5	5	6	37	463602	1010	60374636021010	Block 1010	0.018473	11.822758
6	6	6	37	463601	3009	60374636013009	Block 3009	0.006969	4.460006
7	7	6	37	463601	1007	60374636011007	Block 1007	0.005658	3.621331
8	8	6	37	463700	2020	60374637002020	Block 2020	0.003199	2.047514
9	9	6	37	463800	1004	60374638001004	Block 1004	0.002044	1.30793
10	10	6	37	461902	2038	60374619022038	Block 2038	0.003244	2.07593
11	11	6	37	461902	2035	60374619022035	Block 2035	0.003585	2.294125
12	12	6	37	461902	2036	60374619022036	Block 2036	0.003351	2.144877

0 of 1,736 selected | Filters: | 100%

Figure 9. The original 2010 Census Blocks dataset in .csv format



c2010_Census_Blocks_Pro_Clip

Field: Add Calculate Selection: Select By Attributes Zoom To Switch Clear Delete Copy

	OBJECTID_1 *	Shape *	OBJECTID	STATE	COUNTY	TRACT	BLOCK	GEOID10	NAME	CENSUS_LAN	CENSUS_L_1	Shape_Length	Shape_Area
1	1	Polygon	6	06	037	463601	3009	060374636013009	Block 3009	0.006969	4.460006	541.985857	18039.11967
2	2	Polygon	7	06	037	463601	1007	060374636011007	Block 1007	0.005658	3.621331	487.45325	14646.82273
3	3	Polygon	20	06	037	463500	1006	060374635001006	Block 1006	0.012828	8.210074	741.593315	33207.43878
4	4	Polygon	21	06	037	463601	3005	060374636013005	Block 3005	0.012024	7.69511	723.603604	31124.00756
5	5	Polygon	22	06	037	463601	1005	060374636011005	Block 1005	0.006389	4.089101	298.51733	4322.13408
6	6	Polygon	37	06	037	463400	2000	060374634002000	Block 2000	0.005467	3.498765	482.465076	10850.98003
7	7	Polygon	61	06	037	463500	1012	060374635001012	Block 1012	0.009222	5.902362	623.482382	23872.99405
8	8	Polygon	62	06	037	463500	3016	060374635003016	Block 3016	0.005762	3.687802	548.468549	14916.35022
9	9	Polygon	63	06	037	463601	3008	060374636013008	Block 3008	0.014222	9.101882	785.652539	36814.89024
10	10	Polygon	64	06	037	463500	1001	060374635001001	Block 1001	0.008704	5.570746	529.465812	17141.67525
11	11	Polygon	95	06	037	463500	3005	060374635003005	Block 3005	0.008928	5.714067	619.017733	23112.15956
12	12	Polygon	96	06	037	463500	3003	060374635003003	Block 3003	0.008058	5.157088	719.588472	20858.40964

0 of 139 selected Filters: 100%

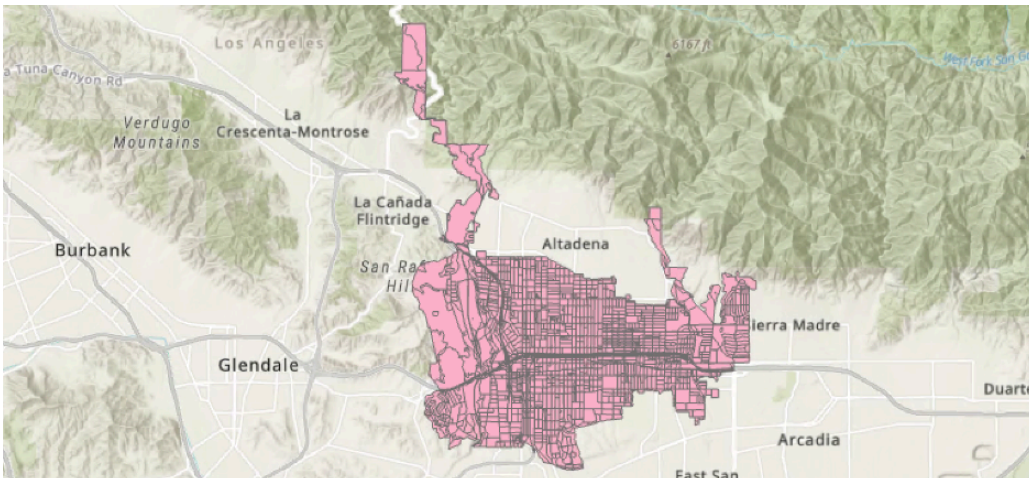


Figure 10. The original 2010 Census Blocks dataset in .shp format

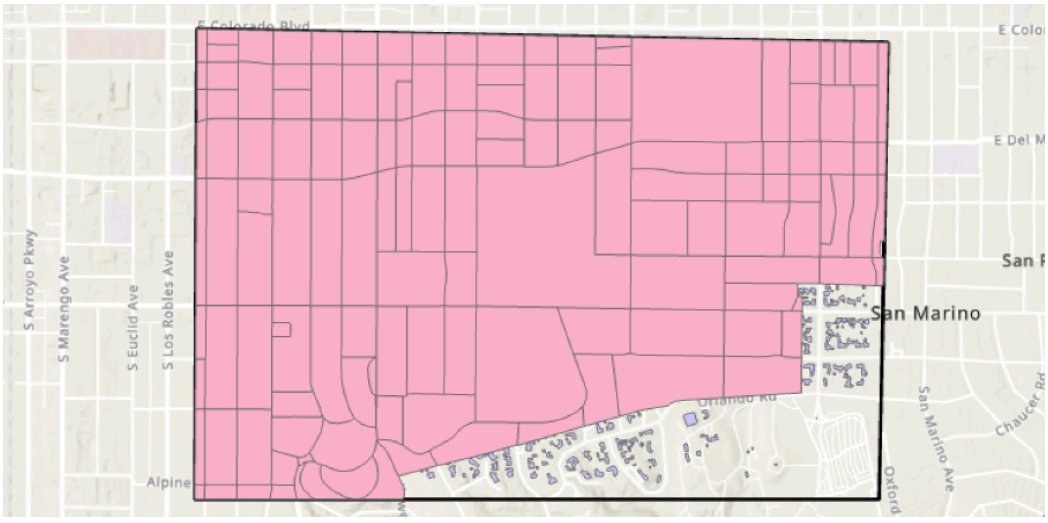


Figure 11. The 2010 Census Blocks dataset after prepared

5.1 Discussion

The methodology adopted for data acquisition and preparation, though rigorous, suggests several enhancements for future endeavors. Specifically, the temporal alignment of datasets could further refine the analysis, considering urban changes over time. Nevertheless, when viewing the prepared datasets collectively, several intriguing patterns emerge. Besides, the proximity of high tree canopy areas to historical buildings raises questions about campus planning and its environmental priorities. Furthermore, the spatial distribution of traffic collisions near pedestrian-heavy zones highlights areas where infrastructure improvements could enhance safety. In conclusion, through this integrative approach, as shown in Figure 12, the assembled spatial data not only furnish a screenshot of the current state but also prepare for longitudinal studies concerning urban development, environmental planning, and community safety within the Caltech campus context.

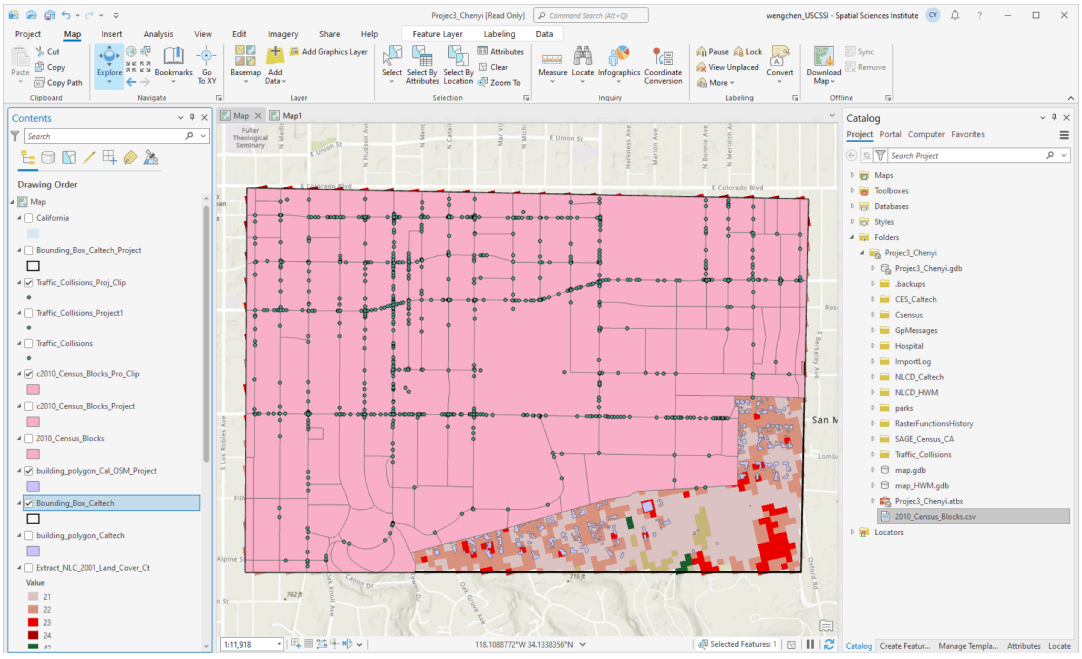


Figure 12. Screenshot relationships between different data types

References

City of Pasadena Open Data Site. "Traffic Collisions Near Caltech Campus," May 5, 2022. https://data.cityofpasadena.net/datasets/85f49ea583c24056968bee6e28162da4_0/explore?location=34.155932%2C-118.127334%2C13.23.

Multi-Resolution Land Characteristics Consortium. "Caltech Building Polygon," 2021. https://www.mrlc.gov/viewer/?downloadBbox=34.06305,34.16221,-118.18258,-117.99883.

Sage Data. "2010 Census Blocks Near Caltech," 2010. https://data-sagepub-com.libproxy2.usc.edu/dataset?view=AA8BXQAAgAC%24AQAAAAAAAAA3_zMslwIJ8Ve1X%24GFlaG5ZTanHOw7iRYJ9HA03_FfoZQ8gfW374m91t18H6Tt_cKdPhe5u18xbimN9SNFtNOp6Sk1e7z6rjM3wpC48lt9kfrDF9W_fm_KPXc2LhNkvz4_pdQON7rWUeq4vmtLa0KWwcaPTcxS1azZmOJTXPZuXZTnqJMed30zuLXmWFmZ7dOuT2MFs9UDegFFyAWMLB4qWuLX%24FDVSCHVNq_jT%24cB4S9cjxuuS1MDyWnAU4VzVAfyrW7MHSJ4tR5Mw.

OpenTreeMap. "Tree Coverage at California Institute of Technology," February 2024. https://www.openstreetmap.org/way/29111188.