Space and Place: How U.S. Public Libraries Evolve with Community Demographics and Needs?

Jiayi Zhang
Data Analytics - School of Arts and Sciences - Tufts University
Sp24-UEP-0235-02-Advanced GIS, Instructor: Sumeeta Srinivasan TA:Laurel Mire

Project Summary

The modern public library began in the 19th century as a facility for storing physical materials and facilitating circulation. Since digital technology and cloud storage had not yet been developed, one of the main functions of public libraries was to satisfy the people's need for information retrieval and ingestion. Planning of Libraries in the social environment in the past were information-oriented. However, in the era of technological exploration, libraries played different roles, corresponding to shifts in people's behaviors and the evolution of society. The project investigates two perspectives on public libraries: human behavior and contemporary society. The first aspect examines how trends in public library usage have evolved over time to assess fluctuations in library visits across different years and predict the growing demand for specific library services. The second aspect involves integrating local demographic information with libraries to leverage the distribution of visitors' characteristics, enhancing library services.

Research Question

- How have the statistics on the status of public libraries in the United States evolved over time, from 2018 to 2022?
- How do public libraries integrate with the local population and leverage the distribution of visitors' characteristics to enhance services based on data from 2021 to 2022?

Literature

Svanhild Aabø, Ragnar Audunson, "Use of library space and the library as place", *Library & Information Science Research*, Volume 34, Issue 2, 2012, Pages 138-149, ISSN 0740-8188, https://doi.org/10.1016/j.lisr.2011.06.002.

Elin Golten, "Public Libraries as Place and Space – New Services, New Visibility", *Recruiting and Managing the New Generation of Employees to Attract New Markets and Create new Services,* Aug. 2019, https://library.ifla.org/id/eprint/2708/1/s09-2019-golten-en.pdf

The two literatures reconsider the role of public libraries and human needs in recent years, highlighting the shift in facility functions to accommodate people's behaviors. Golden's work explores social shifts in Norwegian public libraries, stating that public libraries should be independent meeting places and arenas for public conversation and debate. Aabø's literature proposes a similar possibility for joint activities, including intensive meetings with friends, classmates, or colleagues to collaborate on common assignments or leisure activities. Additionally, programs for children in libraries, most frequently on Saturdays, provide a space where children can behave naturally in ways that might not be allowed in many other learning-oriented areas.

María José Del Barrio-Tellado, Mafalda Gómez-Vega, Jonathan Daniel Gómez-Zapata, Luis César Herrero-Prieto, “Urban public libraries: Performance analysis using dynamic-network-DEA”, *Socio-Economic Planning Sciences*, Volume 74, 2021, 100928, ISSN 0038-0121, https://doi.org/10.1016/j.seps.2020.100928.

This literature examined the performance of certain contextual variables evaluating the efficiency of urban public libraries. It conducted a dynamic network EDA to represent the accessibility for local residents in accessing public library services. The geolocation data, labeled as high-high, high-low, etc., were analyzed along with the minimum, maximum, mean, and standard deviation to scale the efficiency level of public libraries. The results were compared using box plots and kernel density for proximity libraries and park libraries, concluding that proximity libraries are more efficient in library management, while park libraries are more attractive to users.

Data Description

- Public Libraries Survey

The survey data were collected from approximately 9,000 public librarians across around 17,000 individual public library outlets located in 50 states and the District of Columbia in the U.S. It consisted of 182 fields and 9,245 rows, including information on library visits, circulation, size of collections, public service hours, staffing, electronic resources, operating revenues, expenditures, and the number of service outlets. The data was collected since 1988, and the year 2018 to 2022 was used as the basis for the study.

- United State Census Data

The demographic data were obtained from the ACS Census portal. The percentage of population per census block in Cook County and Middlesex County has been used. The data were accessed from the Tufts ArcGIS portal. The final selection for Cook County's Children Population ACS data included 92 rows and 12 fields, featuring the sum and mean of the children's population per polygon and the unique ID for each census block. The final selected row for Middlesex County was 56, associated with corresponding children population percentages and census block IDs.

Method

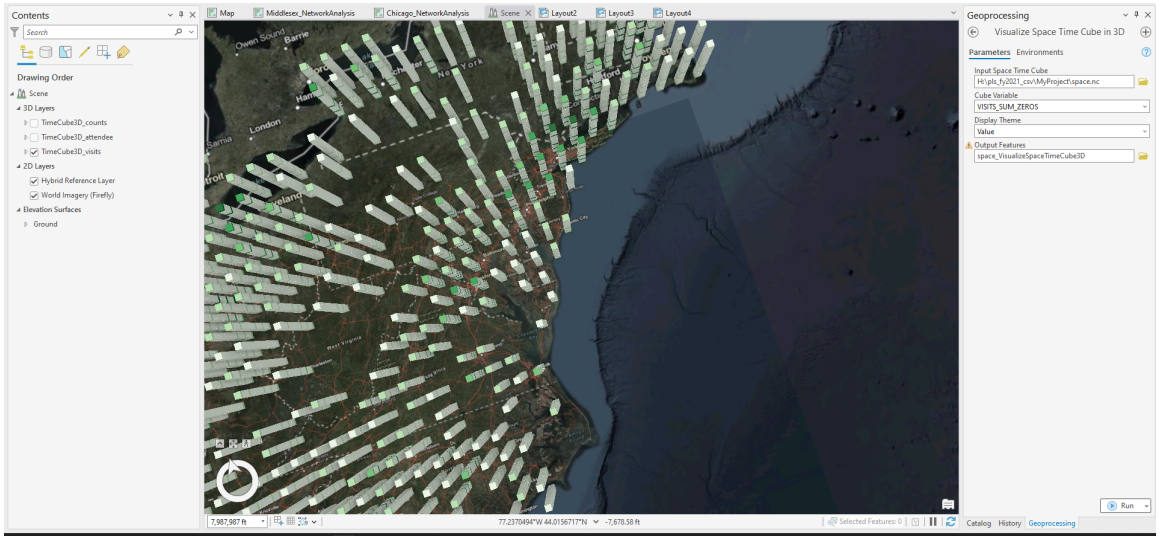
- Spacetime Temporal Analysis

1. This project uses five-year data to observe the trend. The preprocessing for the dataset is to combine them by year. It selects VISITS greater than the mean for each dataset and adds an additional field to represent the YEAR. The five datasets are then merged by the same field, YEAR, completing the data preparation step.

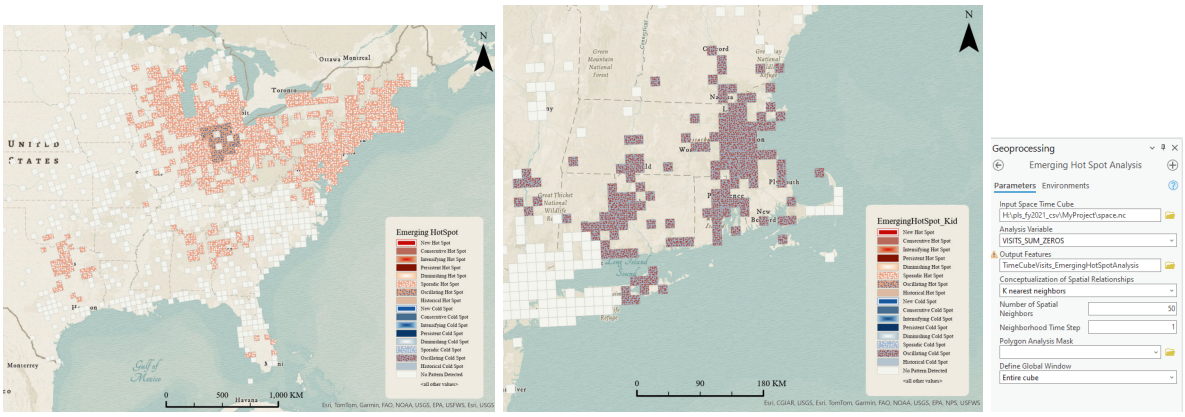
2. To make the prediction, the data points need to be converted into a 2D space-time cube (Space Time Cube by Aggregating Points). Then, they will be visualized in 3D. The emerging hotspot analysis serves as an additional reference to identify the hot and cold spots of VISITS.

3. Finally, place the space-time cube into the forecasting model (Exponential Smoothing Forecast). The forecasting results are shown in the poster.

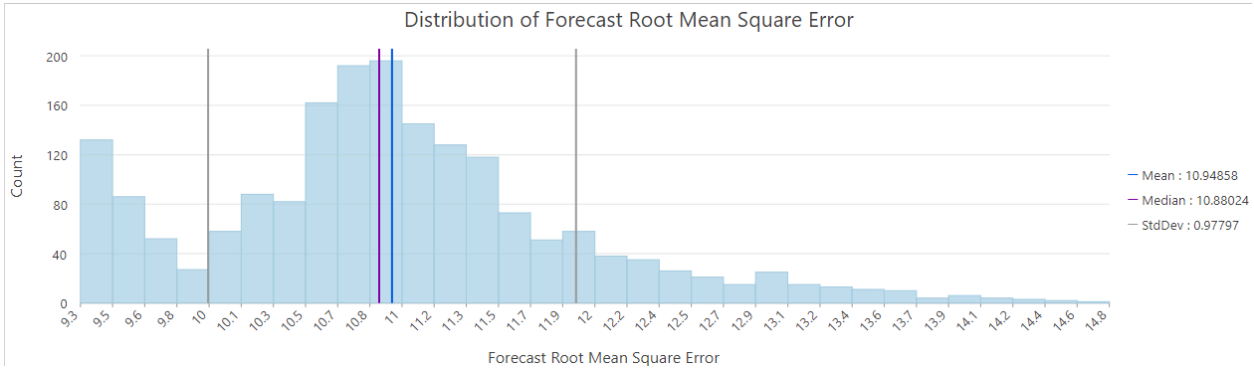
4. The distribution of Root Mean Square Error is approaching a normal distribution within the 97% standard deviation, which indicates that the model's predictions are consistent across the merged datasets. This level of deviation is acceptable and suggests that the model errors are normally distributed, affirming the model's validity under the forecasting operating conditions.



Visualize the spacetime cube 3D by VALUE VISITS



Emerging Hotspot Analysis by VISITS and CHILDREN ATTENDEE



The Distribution of Forcast Root Mean Square Error

- Random Forest Feature Selection

  The prediction type is Train Only, targeting VISITS and CHILDREN ATTENDEE. The number of trees is set to 100, with 70% of the data points for each branch. The validation is processed with 30% of the training data, looping five times. Finally, calculate the mean importance and rank them in descending order by using the result of the best iteration.

	OID	POPU_UND	TOTSTAFF	EBOOK	AUDIO_DL	VIDEO_DL	KIDCIRCL	ELMATCIR	KIDPRO	YAPRO	KIDATTEN	WEBVISIT	BEST_ITE
1	0	1563684655100000	228563818939000	14493516067500	13065178036700	13469992611100	194391913869000	42875716663200	362136444012000	73541389605900	248804615337000	33674737969600	Iterations
2	1	127986726572000	169268179606000	8691741507440	14041157827200	11374794970000	212640904738000	35902421087900	249217749179000	69139915641900	201275102764000	22008015035400	Iterations
3	2	151829544983000	227884844998000	17078520928800	13470886582800	12938706634300	160889208411000	47343308230200	339217864909000	98718636728800	236597010061000	45517500497300	Iterations
4	3	163359492941000	256767331804000	19724979847700	17462612504000	12818114197900	179974619152000	52840432946500	324690985961000	116795111921000	270942648399000	30208616606000	Iterations
5	4	115525469578000	194460350369000	19546276641400	12531190150300	11687087640000	183163972441000	35955043142800	382565698388000	94375938723700	276197769722000	25644714052300	Best Iteration

Distribution of Variable Importance

- Network Analysis

  The network analysis integrates important features from the forecasting results and the local demographic data, specifically the children's population, to create a network analysis of the two most visited libraries: the Chicago Public Library and the Boston Public Library. The libraries in the selected counties act as inputs for the Service Area under the travel model with a driving time cutoff of 10 minutes. After constructing the network, the study summarizes the percentage of the children's population within the network area to identify areas with a higher percentage of children, as well as the libraries included.

Uncertainty and Challenge

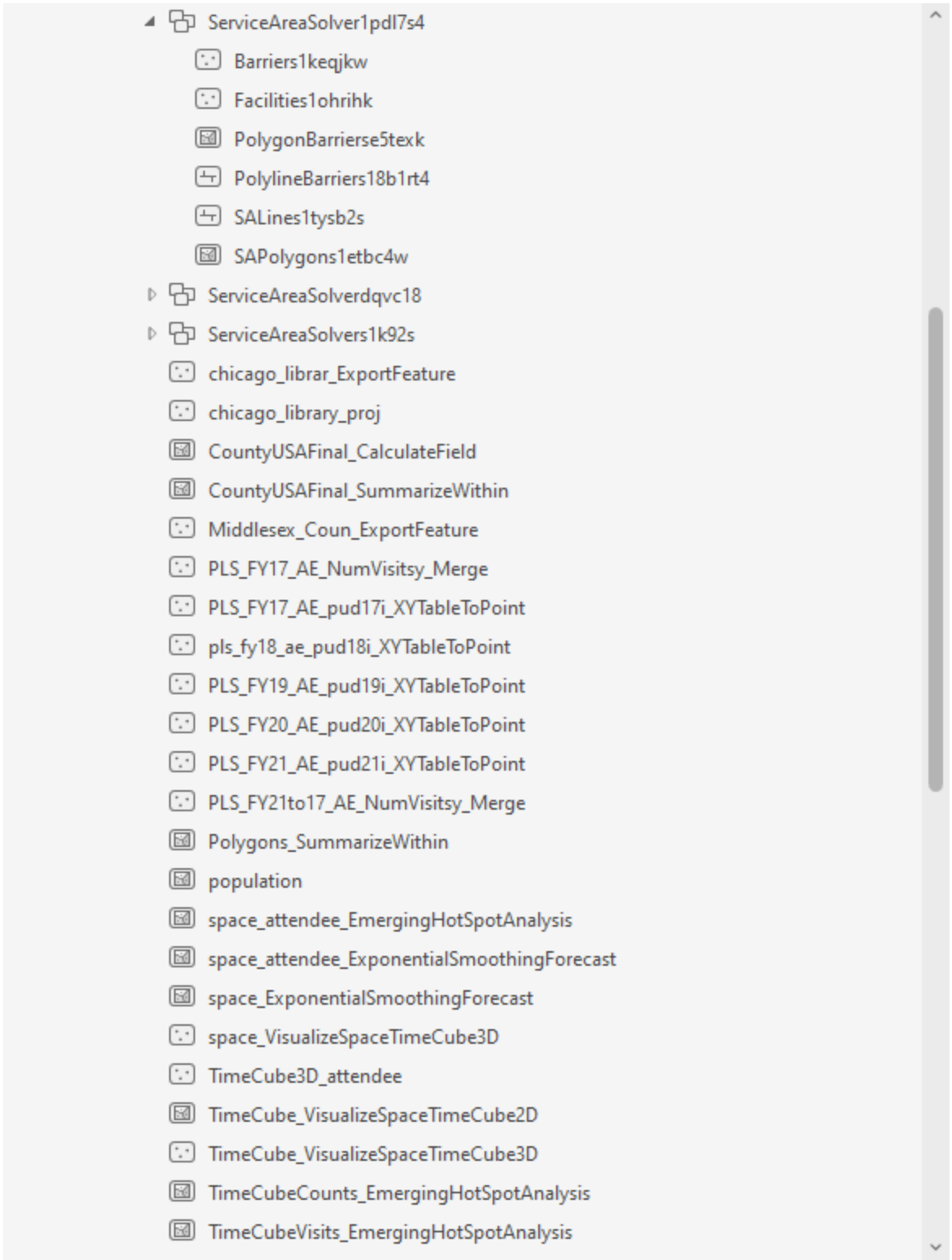
- This project was a follow-up to a previous study that identified public libraries with the highest number of visits in the United States, comparing them with physical item circulation and nearby universities. From that project, it was found that the number of visits was much greater than the number of physical item circulations but consistent with the number of universities around them. This leads to the assumption of how people use public libraries and what motivates them to visit. This project answers this question from one perspective: how does it change with time and what would be the most important features leading to the visits in the coming year. However, it won't be limited to one side Children program. For the continuing work, more variables will be included. On the other hand, from the previous study, the Hawaii Public Library does not have a significant number of universities nearby, yet the number of visits ranks tenth on the list. This is because it is located at the center of attractions and is close to all museums. On Google reviews, most of the reviews mention that the air conditioning there is delightful. Therefore, the nearest attraction is also an important feature that could be analyzed by network analysis.

- The possible error originates from the Forest Base Classification and Regression step. The varying number of branches and validation sets leads to a different order of feature importance. This study attempts several iterations with different branches and validation sets and finds that the Children's program is always among the top 10 most important features. Another possible error comes from the Emerging Hot Spot Analysis. The results vary significantly depending on the conceptualization of spatial relationships. Finally, this study chooses a Fixed Distance of 100 KM instead of K nearest Neighbors subjectively. The lack of a standardized conceptualization of spatial relationships is also the reason for not including emerging spot visualization in the poster.

Conclusion

This project investigates how the function of US public libraries has changed in response to community needs from 2018 to 2022. It forecasts the important features that could lead to changes from 2024 to 2025. Then, it constructs a network analysis based on the selected features. The conclusion is that public libraries with fewer visits in the past might want to consider increasing the number of children's programs to attract more children and their families. The typical cases involving Chicago and Boston public libraries demonstrate that public libraries near these institutions, within the census blocks with a high percentage of children, could also provide more children's programs to share the burden with the Chicago and Boston public libraries, because there are nearly no children living close to these two selected libraries.

Overall, the forecast is acceptable, relying on the large number of data points and articulated geoprocessing steps. However, limitations caused by model parameters and time constraints exist. The sustained development of society leads to unpredictable shifts in future needs, it also opens up possibilities for continually improving the community by responding to changes in data.


ServiceAreaSolver1pdl7s4
Barriers1keqjkw
Facilities1ohrihk
PolygonBarrierse5texk
PolylineBarriers18b1rt4
SALines1tysb2s
SAPolygons1etbc4w
ServiceAreaSolverdqvc18
ServiceAreaSolvers1k92s
chicago_librar_ExportFeature
chicago_library_proj
CountyUSAFinal_CalculateField
CountyUSAFinal_SummarizeWithin
Middlesex_Coun_ExportFeature
PLS_FY17_AE_NumVisitsy_Merge
PLS_FY17_AE_pud17i_XYTableToPoint
pls_fy18_ae_pud18i_XYTableToPoint
PLS_FY19_AE_pud19i_XYTableToPoint
PLS_FY20_AE_pud20i_XYTableToPoint
PLS_FY21_AE_pud21i_XYTableToPoint
PLS_FY21to17_AE_NumVisitsy_Merge
Polygons_SummarizeWithin
population
space_attendee_EmergingHotSpotAnalysis
space_attendee_ExponentialSmoothingForecast
space_ExponentialSmoothingForecast
space_VisualizeSpaceTimeCube3D
TimeCube3D_attendee
TimeCube_VisualizeSpaceTimeCube2D
TimeCube_VisualizeSpaceTimeCube3D
TimeCubeCounts_EmergingHotSpotAnalysis
TimeCubeVisits_EmergingHotSpotAnalysis