Transit Growth and Urban Layout: A Comparative Analysis of Major American and Chinese Cities

By
Gugo Babayan

Project Overview:

This project embarks on a comparative spatial analysis of three major American cities—New York, Los Angeles, and Chicago—and their Chinese counterparts, Beijing, Shanghai, and Guangzhou. My primary research question probes the relationship between transit growth and urban layout changes, hypothesizing that Chinese transit systems lead urban development, whereas in American cities, urban development precedes transit growth.

Research Questions and Motivation:

I sought to understand how transit infrastructure and urban layout changes interact. As an international student, I have always wondered why the American public transportation system is so bad considering that this is the richest country in the world. This is why I wanted to do a comparative study with another country to try to understand how different nations approach the issues of urban planning and public transportation. The question, "What is the relationship between transit growth and urban layout changes in the biggest American and Chinese cities?" is pivotal in understanding urban planning and development strategies across different cultural and political landscapes. In particular, China is now the second biggest economy in the world and has been developing economically at faster rates and the Chinese government has completely changed the transportation infrastructure. This is one of the reasons that I was interested in China, given my background in Political Science.For further reading and background, readers can refer to resources available at [GRC insights] . My hypothesis is that the Chinese transit system leads to urban built up whereas American cities follow urban built up

Data Description and Methodology:

My project includes two primary datasets. The first, sourced from [Citylines.co](https://www.citylines.co/), details transit lines/stations of each city, including geographical attributes and operational timelines. Citylines.co is a collaborative platform for mapping the transit systems of the world. The datasets are open and available under the Open Database License. The second dataset, from the [Atlas of Urban Expansion](http://atlasofurbanexpansion.org/), provides raster data categorizing locations as urban, suburban, rural, or exurban. This dataset is the result of collaboration between NYU, UN-Habitat and the Lincoln Institute of Land Policy.

Data Cleaning and Integration:

I cleaned the datasets by filtering irrelevant data, normalizing different formats, and resolving inconsistencies. Furthermore, I have vectorized the raster data provided by the Atlas in order to make it more accessible as the raster data was very large and complex in structure. The datasets were then merged using Tableau, allowing us to juxtapose the growth of transit systems with urban layout changes over time.

The raster datasets were vectorized by the following method. The raster-to-vector conversion process involves two key steps: resampling the original raster and polygonizing the resampled raster. The *rasterio* library is used to handle raster data, including opening the raster file, resampling it to a specified target resolution, and extracting pixel values. The resampling is performed to reduce computational complexity while retaining essential features. The *geopandas* library is employed to create a *GeoDataFrame*, a tabular data structure with a geometric column containing polygonized features and additional attribute columns. The shapes function from *rasterio.features* is utilized to polygonize the resampled raster, generating a list of features with geometries and associated properties. These features are then converted into a GeoDataFrame for further analysis, visualization, or export. The entire process ensures the conversion of raster data into a vector format suitable for spatial analysis and mapping. You can have a look at the snapshot of the two datasets below.

	id	klass	length	opening	buildstart	closure	lines	geometry
1	656	Section	3641	2008	2005	999999	[{ "line": "Line 8", "line_url_name": "179-line-8", "syst…	LINESTRING (116.3879 39.975…
2	1105	Section	1436	999999	2015	999999	[{ "line": "Capital Airport Express", "line_url_name": "1…	LINESTRING (116.4277 39.940…
3	610	Section	12650	2014	2011	999999	[{ "line": "Line 6", "line_url_name": "177-line-6", "syst…	LINESTRING (116.6091 39.923…
4	681	Section	13639	1971	1965	999999	[{ "line": "Line 1", "line_url_name": "173-batong", "sys…	LINESTRING (116.3458 39.905…
5	613	Section	13683	2014	2010	999999	[{ "line": "Line 14", "line_url_name": "183-line-15", "s…	LINESTRING (116.4727 40.026…
6	616	Section	10616	2008	2008	999999	[{ "line": "BRT 3", "line_url_name": "195-brt-3", "syste…	LINESTRING (116.4062 40.103…
7	618	Section	5498	2012	2007	999999	[{ "line": "Line 9", "line_url_name": "180-line-9", "syst…	LINESTRING (116.3149 39.894…
8	13049	Section	22	2019	2014	999999	[{ "line": "Line 7", "line_url_name": "178-line-7", "syst…	LINESTRING (116.6893 39.861…
9	606	Section	27065	2009	2005	999999	[{ "line": "Line 4", "line_url_name": "175-line-4", "syst…	LINESTRING (116.264 40.0112…
10	612	Section	22549	2014	2009	999999	[{ "line": "Line 7", "line_url_name": "178-line-7", "syst…	LINESTRING (116.315 39.8948…
11	609	Section	29546	2012	2007	999999	[{ "line": "Line 6", "line_url_name": "177-line-6", "syst…	LINESTRING (116.271 39.9313…
12	611	Section	8807	2018	2014	999999	[{ "line": "Line 6", "line_url_name": "177-line-6", "syst…	LINESTRING (116.271 39.9314…
13	677	Section	2292	2017	2014	999999	[{ "line": "Fangshan Line", "line_url_name": "188-fang…	LINESTRING (116.1189 39.721…
14	669	Section	3331	2013	2007	999999	[{ "line": "Line 10", "line_url_name": "181-line-10", "s…	LINESTRING (116.3136 39.843…
15	664	Section	2259	2013	2007	999999	[{ "line": "Line 8", "line_url_name": "179-line-8", "syst…	LINESTRING (116.3891 39.948…
16	662	Section	6171	2013	2011	999999	[{ "line": "Line 8", "line_url_name": "179-line-8", "syst…	LINESTRING (116.3568 40.079…
17	670	Section	5624	2020	2016	999999	[{ "line": "Fangshan Line", "line_url_name": "188-fang…	LINESTRING (116.2995 39.812…
18	678	Section	13453	2017	2014	999999	[{ "line": "Yanfang Line ", "line_url_name": "198-yanfa…	LINESTRING (116.0946 39.727…
19	682	Section	6879	1987	1987	999999	[{ "line": "Line 2", "line_url_name": "174-line-2", "syst…	LINESTRING (116.3508 39.903…
20	685	Section	560	1971	1965	1987	[{ "line": "Line 1", "line_url_name": "173-batong", "sys…	LINESTRING (116.3458 39.905…
21	675	Section	11430	2011	2009	999999	[{ "line": "Line 15", "line_url_name": "184-line-15", "s…	LINESTRING (116.5584 40.112…
22	15800	Section	2129	2021	2014	999999	[{ "line": "Batong Line", "line_url_name": "192-batong…	LINESTRING (116.6895 39.855…

Figure 1: The CityLines data shows the year of the transit line opening ('opening) and if it is still operating ('closure'). It also has descriptions of the transit lines ('lines') such as line name, and system name. Finally, it has the geographical attributes of the line ("geometry") in the format of linestring,type of vector data that consists of a sequence of points, connected by straight line segments.

	raster_val	geometry
1	7	POLYGON ((395136.9 4705952,...
2	5	POLYGON ((395496.5 4705952,...
3	5	POLYGON ((405205 4705593, 4...
4	5	POLYGON ((406643.3 4705952,...
5	5	POLYGON ((410598.6 4705593,...
6	4	POLYGON ((410958.2 4705593,...
7	5	POLYGON ((418149.6 4705593,...
8	6	POLYGON ((420666.6 4705593,...
9	4	POLYGON ((409879.4 4705593,...
10	3	POLYGON ((410239 4705233, 4...
11	3	POLYGON ((410958.2 4705233,...
12	5	POLYGON ((419587.9 4705593,...
13	4	POLYGON ((421026.2 4705233,...
14	3	POLYGON ((421385.8 4705233,...
15	2	POLYGON ((421745.4 4705233,...

Figure 2: This is the second Atlas dataset after cleaning and polygonizing it from the rough raster data format. The 'raster_val' indicates the type of the vectorized pixel in ordered data type from 1 to 7. For example, '1' corresponds to "Urban Built-Up", '2' corresponds to 'Suburban Built-up', '3' to "Rural Built-Up' and so on. The other variable "geometry" is the poligonzied geographic data, a polygon is a closed shape formed by a connected sequence of straight line segments.

Results:

The visualizations reveal distinct patterns in the transit development and urban layout evolution of the selected cities. For instance, the temporo-spatial mapping of American cities illustrates that the transit lines are built in highly urbanized places whereas the Chinese cities tend to first build transit which is followed by urbanization. However, there are some outliers such as the NYC which started its extensive transit lines almost 2 centuries ago and throughout this time it has built transit lines in non-urban places too, thus rejecting the hypothesis. You can see pictures of the fina visualizations below.

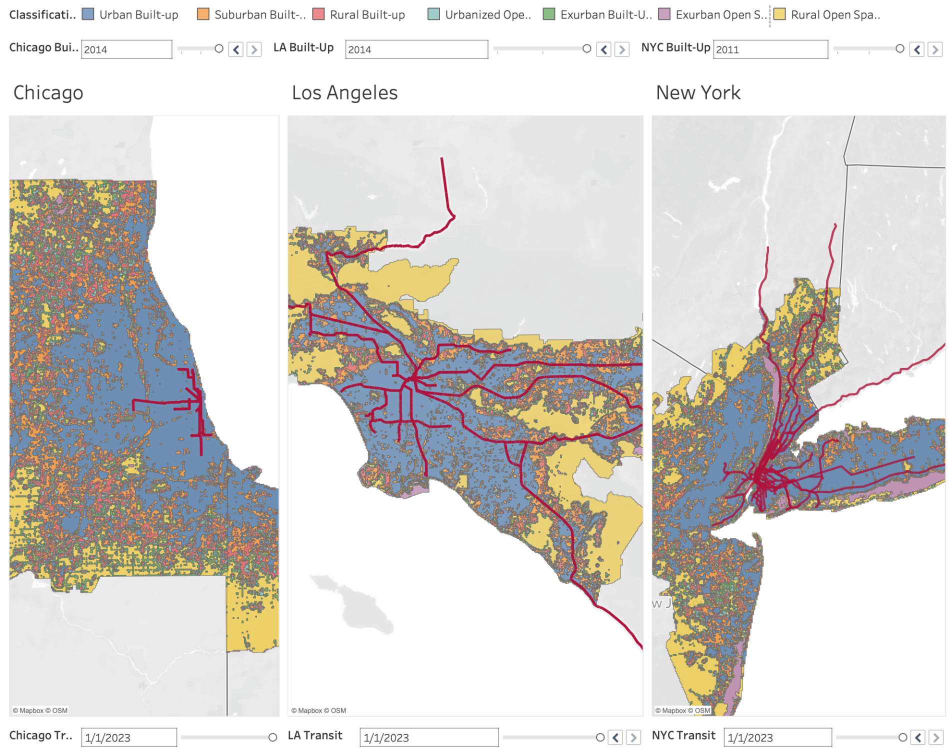

Figure 3: This visualization shows the transit lines and the urban layout patterns in three biggest American cities. These maps are color-coded to represent different types of land use and the legend explaining the color code is at the top. There are also sliders with dates at the top and at the bottom of each city. The top slider relates to the built up patterns and when the user changes the year the built up changes accordingly. The bottom slider relates to the transit lines, if the user interacts with the year the transit lines change on the map accordingly.

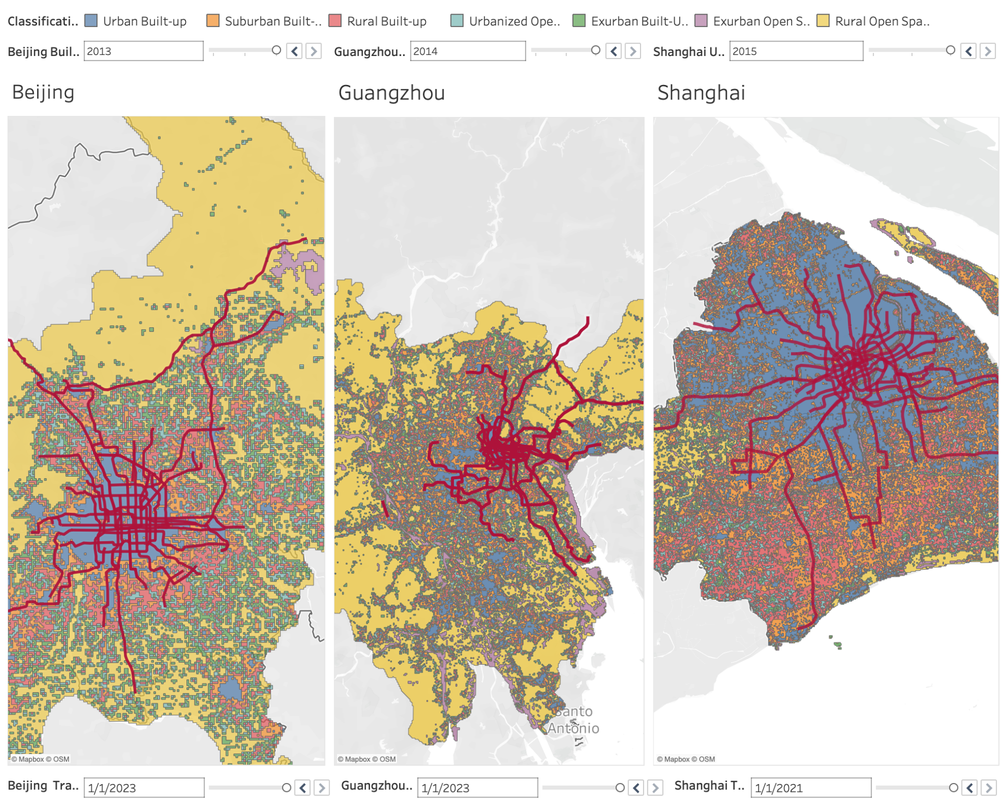

Figure 4: This visualization shows the transit lines and the urban layout patterns in three biggest Chinese cities. These maps are color-coded to represent different types of land use and the legend explaining the color code is at the top. There are also sliders with dates at the top and at the bottom of each city. The top slider relates to the built up patterns and when the user changes the year the built up changes accordingly. The bottom slider relates to the transit lines, if the user interacts with the year the transit lines change on the map accordingly.

In comparing the two, it seems the Chinese cities have a more extensive reach of transit lines beyond the urban core, potentially indicating a more aggressive approach to connecting outlying regions with the urban centers. On the other hand, American cities show a concentration of transit lines within the urban core with less penetration into the suburban and rural areas, which may reflect the car-centric development patterns historically prevalent in the United States. These patterns can be indicative of differing urban planning philosophies, with Chinese cities possibly focusing on public transit as a means to manage urban sprawl and connect different urban and rural areas. American cities may exhibit a pattern where public transit is more focused on serving the dense urban core, with other areas more reliant on personal vehicles for transportation

Furthemore, I have demonstrated the growth of the transit lines and the changes in the urban layout in line graphs and bar charts respectively.

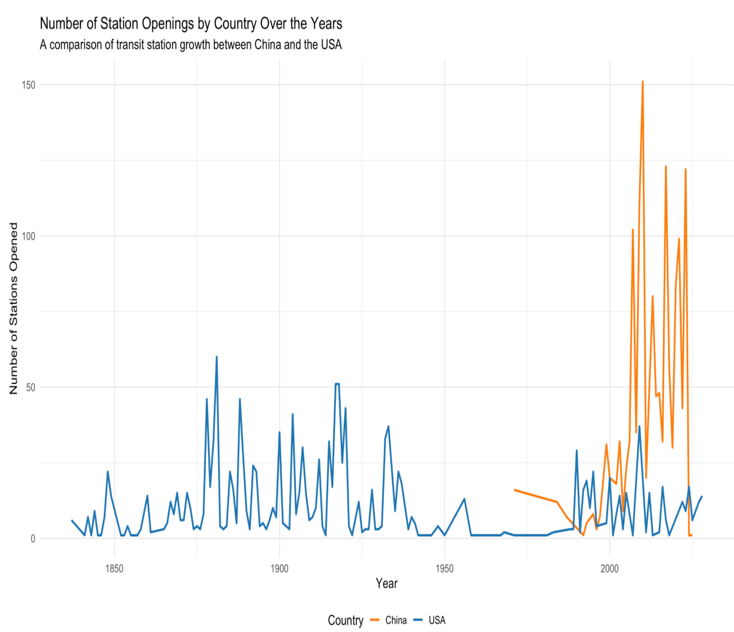

Figure 5:

The line graph contrasts the transit station development in China and the USA from 1850 to the present. The USA's growth is modest with a downward trend in the late 20th century, depicted by the blue line. China's transit development, shown in orange, is minimal until an explosive rise beginning in the 21st century, indicating a recent, massive expansion in its public transit infrastructure.

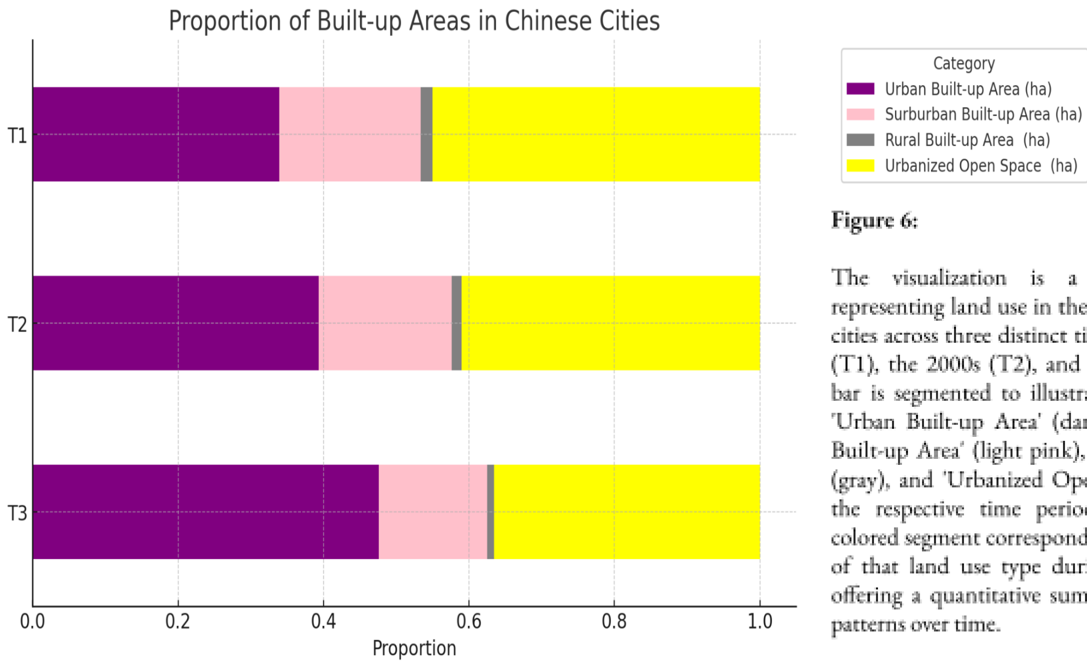

Figure 6:

The visualization is a stacked bar chart representing land use in the mentioned 3 Chinese cities across three distinct time periods: the 1990s (T1), the 2000s (T2), and the 2010s (T3). Each bar is segmented to illustrate the proportion of 'Urban Built-up Area' (dark purple), 'Suburban Built-up Area' (light pink), 'Rural Built-up Area' (gray), and 'Urbanized Open Space' (yellow) for the respective time period. The size of each colored segment corresponds to the relative extent of that land use type during the given decade, offering a quantitative summary of urbanization patterns over time.

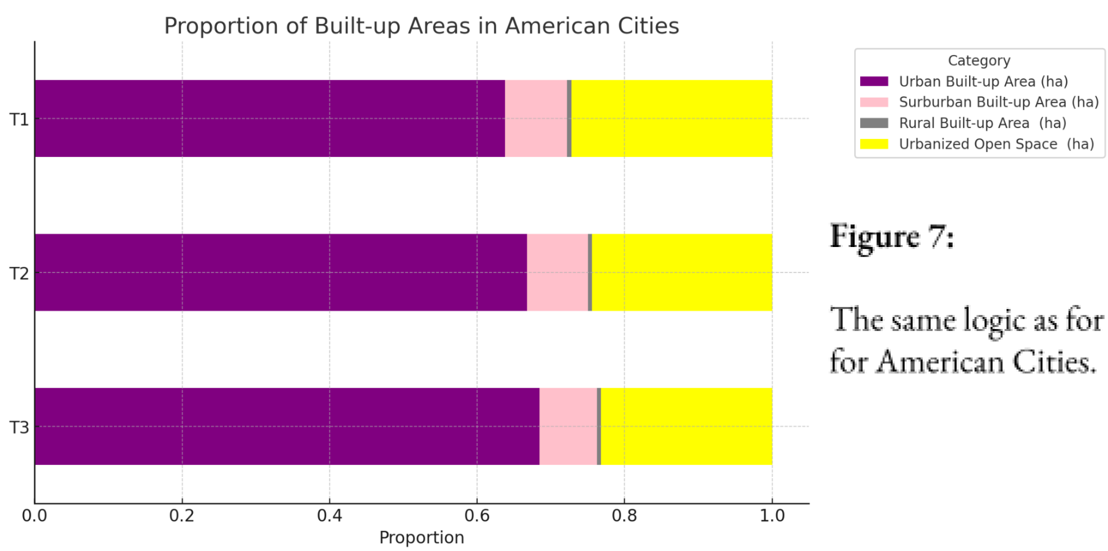

Figure 7:

The same logic as for *Figure 6* above but for American Cities.

In comparing the two countries' megacities built up proportions, we can see that the American cities have significantly higher urban built up area with little urbanized open space. In contrast, there is a more balanced view for the Chinese cities where the Urban built up is growing (see the difference between T1 and T3) but there is still significant urbanized open space too. In the context of urbanization, Chinese cities appear to have maintained a consistent pattern of suburban growth, while American cities have experienced significant urban consolidation. As for transit growth, the American pattern suggests a potential underinvestment in transit infrastructure in suburban and rural areas, while the Chinese pattern implies aggressive expansion and integration of transit services, in line with the increasing urban density.

Open Questions and Limitations:

In assessing the relationship between transit growth and urban layout changes in the largest American and Chinese cities, several key questions remain open. Determining whether transit infrastructure development is a driving force behind urban buildup or simply correlates with it due to other variables like economic trends or population growth is essential. The influence of urban planning policies, historical and cultural contexts, and their impact on socioeconomic factors also remain areas to be explored. Furthermore, understanding how these patterns might inform future urbanization and transit trends is crucial.

The research is not without limitations, starting with potential issues in data availability and the complexity of factors influencing urban dynamics that may overshadow the role of transit growth. The timeframe of the study may not capture the full scope of the evolutionary trends, and regional variations could challenge the generalizability of findings. In particular, the urban layout datasets encompassed only 3 historical periods (90s, 2000s, and 2015s) making it difficult to thoroughly understand changes in layout before and after a line is opened. This data limitation and the time constraints made it impossible for me to visualize a more clearer non-spatial graph that explicitly shows the before-after effects of transportation on urban layout. Instead, one can see this by only interacting with the maps and year sliders. This might be an interesting topic to explore in the future to perhaps cluster the transit lines based on urban, suburban, rural locations. Furthermore, it might have been better to include more cities in the sample rather than the 3 biggest cities of each country but, again, there was unfortunately no transit data on other cities.

Reproducibility and Resources:

The project' reproducibility is ensured through an organized repository of RMD(where you can see the code behind the non-spatial visualization) and a Tableau file in .twbx format which includes the datasets and makes everything accessible to the viewer.

Conclusion:

This project's exploration into the intricate dance between transit growth and urban layouts in major American and Chinese cities concludes with nuanced findings. The comparative analysis, leveraging temporal and spatial data, reveals that Chinese cities exhibit a pattern where transit infrastructure potentially catalyzes urban development. In contrast, American urban expansion does not consistently precede transit growth; rather, transit development seems to follow or coincide with urbanization, with the notable exception of New York City, which exhibits characteristics of both models. This suggests a complex relationship where infrastructure and urban growth can be both cause and effect, influenced by a myriad of factors beyond a simple linear progression.

This project has also been an opportunity for me to think more about the relationship between data and story. I have been constantly in contact with experts on the field who helped me to pivot a lot during challenging times in the project without giving up on it. I have learned a lot not only about transit systems but about how to approach a data science project, how to conduct interviews with field experts, and how to present findings to a general audience. Furthermore, I learned a lot from working with my classmate, Bryan, as we have been navigating through the challenges of the projects when it came to understanding how to divide work, how to keep the interconnectability of the two projects, and how to present both of our projects in a single format.