Title

Mass Transit Stop and Route Inventory and Mapping: Development and Refinement of a Protocol

Author

Mills, Jackson

Publication Date

2025-07-01

UC Berkeley
Institute of
Transportation Studies

Mass Transit Stop and Route Inventory and Mapping: Development and Refinement of a Protocol

Author Jackson Mills, University of California, Berkeley

Advisor Daniel Rodríguez, University of California, Berkeley

July 2025

Pacific Southwest Region UTC
University Transportation Center

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.
UCB-ITS-PSR-2025-06</td><td>2. Government Accession No.
N/A</td><td colspan="2">3. Recipient's Catalog No.
N/A</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Mass Transit Stop and Route Inventory and Mapping:
Development and Refinement of a Protocol</td><td colspan="2">5. Report Date
July 2025</td></tr>
<tr><td colspan="2">6. Performing Organization Code:
ITS Berkeley</td></tr>
<tr><td colspan="2">7. Author(s)
Jackson Mills, M.C.P.</td><td colspan="2">8. Performing Organization Report No.
N/A</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Institute of Transportation Studies (ITS Berkeley)
University of California, Berkeley
109 McLaughlin Hall MC1720
Berkeley CA 94720-1720</td><td colspan="2">10. Work Unit No.
N/A</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
USDOT Grant 69A3551747109</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
1200 New Jersey Avenue, SE, Washington, DC 20590</td><td colspan="2">13. Type of Report and Period
Whitepaper (August 2023 - May 2024)</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
USDOT OST-R</td></tr>
<tr><td colspan="4">15. Supplementary Notes</td></tr>
<tr><td colspan="4">16. Abstract
This report details the process of geospatially mapping every heavy rail, light rail, and bus rapid transit expansion in the United States from the years 2000 to 2024. We outline the protocol that was developed for this project and explain the steps that were taken to produce route and stop shapefiles for 148 transit openings and extensions. We address some of the challenges we encountered. We also produce a series of visualizations to illustrate trends in the geographic and modal distribution of transit projects in the United States over the last 25 years. In mapping these expansions, we hope to provide researchers with the requisite information to be able to conduct a wide range of studies that examine multiple types of effects associated with public transit on a wider scale. Such analysis could be conducted on a transit line- or stop-based level, which are the two shapefiles produced for each transit extension in this project. The shapefiles could be modified with catchment areas to examine a transit system's impact within a specified geographic area around the transit line or stop(s). Temporally, researchers could analyze the effects of transit expansion on various co-benefits by comparing outcomes before and after the extension opened. By providing the geospatial data, sourcing, and explanations for research practices used in this project, our work serves as a foundation for many types of studies that examine public transit-related outcomes.</td></tr>
<tr><td colspan="2">17. Key Words
Digital mapping, GIS, Public transit, Fixed routes, Transit stops, Transit schedules, Transit data</td><td colspan="2">18. Distribution Statement
No restrictions</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
46</td><td>22. Price
N/A</td></tr>
</table>

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes the following partners: California State University, Long Beach, Northern Arizona University, Pima Community College, University of California, Berkeley, University of California, Davis, University of California, Irvine, University of California, Los Angeles, University of Hawaii, and University of Nevada, Las Vegas.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* The program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

US Department of Transportation Disclaimer

Disclosure

This research took place at the University of California, Berkeley from October 2024 to June 2025 and was supported by a grant under the Pacific Southwest Region University Transportation Center research program in the amount of $7,500.

Table of Contents

Executive Summary

This report details the process of geospatially mapping every heavy rail, light rail, and bus rapid transit expansion in the United States from the years 2000 to 2024. We explain some of the possible uses of this work in future research, from engaging with transit's role in public health outcomes to analyzing its land use and socioeconomic impacts. We outline the protocol that was developed for this project and explain the steps that were taken to produce route and stop shapefiles for 148 transit openings and extensions. We address some of the challenges we encountered. Lastly, we provide a series of visualizations to illustrate trends in the geographic and modal distribution of transit projects in the United States over the last 25 years.

1. Context and Objectives

Mass transit systems can carry a wide range of benefits for cities and regions, ranging from providing viable alternatives to driving to reducing congestion and travel times (Beaudoin et al., 2015, Aftabuzzaman et al., 2010). There are many additional co-benefits of public transit, such as outcomes related to public health in terms of physical activity, pollution reduction, and road safety (Gendron-Carrier et al., 2022, Kim et al., 2022, Li et al., 2019). Public transit use can help address high rates of physical inactivity by encouraging higher rates of active transportation (i.e., walking and cycling) to access transit (Xiao et al., 2019). Transit may contribute to safer road conditions and fewer road accidents (Albalate and Fageda, 2019). There are fiscal benefits from transit, ranging from economic investment in previously neglected communities to increased tax revenues in well-served areas (Weisbrod and Reno, 2009). Transit can have significant effects on the surrounding built environment, transforming land use to be more complementary for transit utilization (Hu et al., 2016, Hurst and West, 2014). These changes, in turn, can influence localized and regional population growth, while also raising concerns about displacement in areas experiencing rapid shifts in urban form (Delmelle, 2021).

This project lays the groundwork for studying the outcomes and co-benefits of public transit growth by geospatially mapping every public transit expansion project in the United States from the years 2000 to 2024. Openings and extensions of heavy rail, light rail, and bus rapid transit systems are mapped. Previous studies have typically focused on no more than a handful of transit extensions when examining various co-benefits; this work gives researchers access to geospatial data for over 140 transit extensions across three mode types. In mapping these expansions, we hope to provide researchers with the requisite information to be able to conduct a wide range of studies that examine multiple types of effects associated with public transit on a wider scale. Such analysis could be conducted on a transit line- or stop-based level, which are the two shapefiles produced for each transit extension in this project. The shapefiles could be modified with catchment areas to examine a transit system's impact within a specified geographic area around the transit line or stop(s). Temporally, researchers could analyze the effects of transit expansion on various co-benefits by comparing outcomes before and after the extension opened. By providing the geospatial data, sourcing, and explanations for research practices used in this project, our work serves as a foundation for many types of studies that examine public transit-related outcomes.

Our main objectives for the project were the following:

1. Develop, test, and refine a protocol for identifying transit extensions to include in the analysis.
2. Assemble a transit opening/extension database, complete with sourcing.

3. Locate, download, and refine transit maps using geographic information systems (GIS).
4. Summarize and visualize key findings using maps and data visualization techniques.

2. Protocol

There were a few questions and challenges associated with the initial planning stages of producing transit expansion geospatial data. We had to identify the best method for locating every transit opening and extension in the United States over the last 25 years. We considered the possible main sources of information and mapping for each expansion. We also had to classify each extension by mode type (heavy rail, light rail, or bus rapid transit), which necessitated the creation of a benchmark that would include or exclude transit extensions based on a consistent set of criteria.

2.1. Initial Steps

The first step was to decide what could be considered transit opening or extension. For the purposes of this research, a transit opening is defined as the launch (or re-launch, in rare cases) of a new public transit service that previously did not exist. Openings typically include both stops and right-of-way (lines); infill stations (stations that open somewhere along an existing right-of-way) are openings with an individual stop. Extensions are defined as an opening that extends an existing line's right-of-way. Both fall under the term "expansion".

We also considered what types of transit modes to include. We turned to the National Transit Database (NTD) of the Federal Transit Administration (FTA) glossary for guidance on the multiple classifications for different modes of transit. The glossary contains definitions for heavy rail (HRT), light rail (LRT), and bus rapid transit (BRT) mode types. (Refer to *Appendix A*

– *National Transit Database Definitions*, for these definitions.) The NTD classifies individual transit services by mode type, contained within each listed transit agency. For example, if a transit agency operates a light rail and a bus rapid transit system, both would be noted under said agency in the database. We used this information on transit mode types to locate every heavy rail, light rail, and bus rapid transit system in the U.S. (Not every transit service that was classified as one of these three modes was included; for more information, refer to Point 1.8 in *Appendix B – Data Practices*.) We focused on these three mode types because they are the highest-capacity transit services in the United States and have the most readily available information on expansion, as well as their benefits and co-benefits.

After locating each transit service, we had to create a database containing information on each system's expansions over the last 25 years. We chose to focus on system expansions only from the years 2000 to 2024 for two main reasons: 1) Information on transit expansions before the year 2000 can sometimes be difficult to procure—such as the total number of stops or the precise date of opening—and 2) the public health-related research that will be conducted using these transit shapefiles will rely on data points (i.e., emissions/pollution data) that would be more difficult to obtain for extensions completed in the 20th century.

We searched for information on every HRT, LRT, and BRT services' expansions with a variety of sources: Transit agency websites often contained historical information or a timeline for when their system expanded in recent decades. The FTA publishes project documents for many expansions that were the recipients of federal grant funding; these documents usually had information such as projected opening dates, the number of stations, track length, and more. Each expansion was compiled into a database that included each opening/extension, the city or region that the project was located, the year that the project opened or was extended, and sourcing.

2.2. Creating the Transit Maps

The next step upon completion of the database was to begin the process of mapping each transit expansion. We turned to Transitland, an open-source data platform that contains geospatial information for more than 2,500 transit agencies across 55 countries (Transitland, 2025). Transitland is powered by General Transit Feed Specification (GTFS) data provided by transit agencies and contains GeoJSON files for most public transit lines in the United States. Each line has JSON files for the routes and stops that are available to download, as well as stop buffers and stop hulls. We downloaded the route and stop files and organized them into folders created for each transit agency. Every JSON file used for this project originated from Transitland, with the exception of a few Los Angeles Metro light rail expansions and portions of New Jersey Transit's Hudson-Bergen light rail system. The previous and/or current iterations of these systems were unavailable on Transitland, necessitating alternative sourcing for these two agencies. More information can be found under Point 2.5 in *Appendix B – Data Practices*.

The route and stop GeoJSON files were imported into ArcGIS Pro and converted to feature classes for editing. Transitland's raw JSON files include the full extent of currently operational transit lines and do not have historical data for transit expansions. This impediment required us to clip the transit lines so that only the stops and route sections associated with each expansion were included in each shapefile. The feature classes were also projected into the USA Contiguous Albers Equal Area Conic (EPSG:102003) projected coordinate system. This map projection offers strong distance accuracy across the contiguous United States. The two exceptions were the HART Skyline HRT opening in Honolulu, Hawaii, and the Tren Urbano

HRT opening in San Juan, Puerto Rico, which were respectively projected to NAD83 / Hawaii zone 2 (EPSG:26962) and 1983 StatePlane Puerto Rico Virgin Islands (EPSG:32161). Once reprojected and clipped, the line and stop files were both saved as separate shapefiles. This gave each extension two shapefiles: One for the line data, and one for the stop data.

2.3. The Exported Shapefiles

Once exported, the shapefiles' attribute tables needed significant modifications and additions to provide useful information about each expansion project. We directly modified each shapefile's associated database file (.dbf), which is linked to the shapefile and will appear when the attribute table is opened in GIS software or when opened separately in another application. The line files contain the following data columns:

- *line_id*: A unique identifier for each opening/extension. The corresponding line_id is also included in the stop tables, thereby linking the two tables.
- *agency*: The name of the transit agency responsible for the operation of the opening/extension. Also includes an Agency ID and Onestop ID, both from Transitland. An Agency ID identifies a transit agency within a GTFS feed. A Onestop ID is an alphanumeric, global, immutable identifier for transit agencies.
- *headway*: The time interval between consecutive vehicles on the route, expressed in minutes.
- *stop_count*: The total number of stops with unique names for the opening/extension. More information on how this is calculated can be found in the "Data Practices" document.
- *length_km*: The total one-way length, in kilometers, of each opening/extension. Calculated in ArcGIS Pro.

- *fuel_type*: The fuel type used on each opening/extension. Only included for bus rapid transit systems.

The stop files contain the following data columns:

- *line_id*: A unique identifier for each opening/extension that corresponds to the line ID displayed in the line tables, linking the two together.
- *stop_id*: A station identifier, provided by transit agencies and included in the original JSON file from Transitland. Not unique.
- *stop_name*: The name of each transit stop. Included in the original JSON file from Transitland.
- *latitude*: The latitudinal coordinate provided for each transit stop under *stop_name*.
- *longitude*: The longitudinal coordinate provided for each transit stop under *stop_name*.
- *year_open*: The year the station opened.
- *headway*: The time interval between consecutive vehicles on the route, expressed in minutes. Only included in the stop file for infill stations; otherwise this variable is included in the line file.

Each expansion's two shapefiles have the same naming convention, differentiated with "Line" or "Stops" at the end of the title. For example, for Bay Area Rapid Transit's SFO Airport extension in 2003, the two files are named "BART_SFO_Airport_Extension_2003_Line.shp" and "BART_SFO_Airport_Extension_2003_Stops.shp". They are both located within a folder titled, "BART_SFO_Airport_Extension_2003". Each opening/extension is contained within a folder for each transit agency. A visual example of the two attribute tables for the line and stop files (as they appear in ArcGIS Pro) can be viewed in *Appendix C – Miscellaneous Visuals*.

2.4. Sourcing

Sourcing was required for information on openings/extensions, schedules (for headway information), and fuel type (for noting which BRT systems operated with fossil fuels and those that used battery-electric buses). As discussed above, transit agency websites and FTA documents were the primary sources of information for system expansion projects. Timetables and schedules were also sourced directly from transit agencies. The fuel type was recorded from a variety of sources ranging from news articles, transit agency websites, personal observation, and federal, state, and local documents.

3. Managing Challenges

There were several dilemmas we encountered while searching for, downloading, editing, and saving the final product of this research. This section will explore three such cases and describe the thought processes that went into resolving them. These challenges are also explained in *Appendix B – Data Practices*, in addition to every additional situation we faced and how we resolved all project-related issues.

3.1. Extension Starting Points

The first case involved deciding where an extension should start on the map. There were two common situations: An extension was either built directly off of an existing station, or shared track with an older section of the right-of-way (ROW) beyond an existing station before branching off onto a new ROW further down the line. Two examples of these situations are found in San Francisco, California, and Portland, Oregon. In San Francisco, Bay Area Rapid Transit's aforementioned SFO Airport extension was built directly off of the existing Colma station, which opened in 1996. In Portland, TriMet's opening of its MAX Green Line in 2009 saw the line share existing tracks with the Blue Line (opened in 1986) before branching off on its own ROW some distance beyond the older station, Gateway Transit Center. We faced a decision as to whether to include the older station (that is not a part of the extension) in the stop shapefile, or clip the line so that the station would not be included and would fall just outside the boundaries of the extension.

We decided on the following solution: For extensions that are built directly off of an existing station with new ROW immediately beyond the station, the older stop that is not a part of the extension would be included in the stop shapefile. This would also add the older station to the attribute table. We elected to practice this method in order to visually demonstrate how an extension connects to existing infrastructure. The older station is easily identified in the attribute table because its opening year is almost always different from the rest of the stations that are a part of the extension. For example, Colma is listed in the table as having opened in 1996, while every other station along the SFO Airport extension is marked as opening in 2003.

For extensions that share an older ROW some distance beyond an existing station, we elected to cut off the extension at the point where it splits from the older ROW so that only new ROW is included in the line shapefiles. No older stations are included in this situation as a result. We decided on this practice because we did not want to include any older ROW along any portion of the line file for extensions. We also wanted to give a visual cue that an extension shares an older ROW by not including an older station from which the extension starts. While some extensions that branch off from older ROW may only run on older tracks for a short distance, there are also extensions that run on older tracks for miles beyond an existing station before branching off. We could not justify including the older ROW in these cases, and it made sense to extend this reasoning to the use of every older track section regardless of length. In the case of TriMet's MAX Green Line opening, the line shapefile only includes the ROW up to the point where it intersects with the older track that the Blue Line uses, and does not include Gateway Transit Center station.

3.2. Duplicate Stops

It was not uncommon for the raw GeoJSON files from Transitland to have duplicate stop names. Many of the files contained bi-directional line and stop data, meaning that there were two overlapping transit lines representing both directions of travel. There were also two listed stations for many individual stops. Some stops would be listed twice with no textual differences in the entries between the two rows; others would have slight variations, such as both rows representing the same stop containing directional information (e.g.,"eastbound" and "westbound"). The existence of duplicate stop names presented a challenge for counting the total number of stops.

We decided to retain the duplicate stops in the stop shapefiles because of the difficulty in determining whether a station was truly two separate stops or simply had two stop markers representing bi-directional service. Many stations—particularly on LRT and BRT systems—did have separate platforms in close proximity to each other. Some stations shared the same name but their platforms were separated by an intersection. Others were a block apart and had slightly diverging names. Keeping every stop in the shapefile was important because many LRT and BRT lines run on separate ROWs for different directions. LRT trains may run on two one-way streets that are one block away from each other, for example. It would be confusing if we deleted some stops that represent the same station but kept the stops of stations that were slightly separated. We would be removing stops for one of the directions that the line operates, which would appear idiosyncratic for viewers of the stop file attribute tables. Some stations would have duplicate names while others would not. We determined that leaving the stations in the stop files alone was the most optimal choice.

Not wanting to count duplicate stop names twice, we devised a multi-step strategy to get the count of the number of stops per extension as accurate and precise as possible. Many expansions had only a handful of stops and could be counted manually. For openings and extensions with many more stops, we first used a formula within each stop shapefile that counts the number of unique row entries and ignores duplicates. But this formula would still count two stations that had identical names but included additional text, such as “eastbound” or “westbound”. To resolve this issue, we created a master stop file that contained every station from every expansion project. We cleaned this file so that all duplicate stations were removed regardless of naming convention. We counted the number of stops per expansion and cross-referenced this figure with that which was generated from the first formula. Discrepancies were examined and the stop counts corrected. We completed a final manual check through the cleaned master stop file to ensure that there were no duplicates and all stations were accounted for.

3.3. Headways

A question arose early on as to how we should record headways (the time interval between vehicles on a specific route). Timetables are subject to change frequently for many transit lines. Headways can be irregular. Some ROWs have multiple lines with different headways. Service frequency also varies depending on the time of day and type of day. We debated whether to report the headways of a transit service when it first opened or was extended, or record systems’ headways from their current service levels.

We concluded that reporting present-day weekday peak-hour (morning and evening commute) headways in the line files' attribute data was the most optimal solution. Finding older timetables from when transit agencies began revenue service on an expansion proved challenging. Information was often missing or time-consuming to locate. With many agencies adopting new post-pandemic timetables and/or struggling to restore pre-pandemic service headways, we posited that recording current headways from the time of day that agencies typically run their most frequent service would be at least somewhat representative of a historical "average" headway for transit services that opened before the pandemic.

Some expansions' ROWs were shared by more than one transit line, complicating the matter of recording headways. Some ROWs contained multiple lines that had the same headway which combined to make service more frequent along the corridor, while others were served by many lines with vastly different headways. Bay Area Rapid Transit's North San Jose extension is an example of the former: With shared service between Orange and Green lines, which individually operate at 20-minute intervals, the lines combine to provide effective 10-minute headways during peak hours. An example of the latter circumstance is Hartford's CTFastrak BRT corridor, which is used by more than ten bus routes. These routes have diverging headways ranging from hourly to every 7.5 minutes.

We prescribed different solutions for these two situations: In the case of an extension like BART's with two routes whose headways combine to produce a more frequent service, we recorded the combined headway in the line file. The headway for the North San Jose extension is listed as 10 minutes. In Hartford's case, it would be challenging to combine the headways of every route given the sheer number of schedules and the irregularity in headways. For situations like these, we took the headway of the most frequent line and reported it as the headway for the expansion.

4. Summary Statistics and Visualizations

Table 1

Summary Statistics															
5-Year Period	HRT: # of Openings	HRT: # of Stations	HRT: Km Opened	HRT: Metro Areas	LRT: # of Openings	LRT: # of Stations	LRT: Km Opened	LRT: Metro Areas	BRT: # of Openings	BRT: # of Stations	BRT: Km Opened	BRT: Metro Areas	Total: # of Openings	Total: # of Stations	Total: Km Opened
2000-2004	7	32	68.43	5	26	189	298.05	15	1	4	1.80	1	34	225	368.28
2005-2009	0	0	0.00	0	21	151	198.89	16	5	123	57.34	4	26	274	256.23
2010-2014	3	7	21.20	3	19	108	186.37	11	7	106	95.80	7	29	221	303.37
2015-2019	4	7	26.84	2	18	83	126.15	8	9	192	165.64	7	31	282	318.63
2020-2024	5	19	50.57	4	14	55	85.28	9	9	187	126.22	9	28	261	262.07
Total	**19**	**65**	**167.04**	**14**	**98**	**586**	**894.75**	**59**	**31**	**612**	**446.80**	**28**	**148**	**1263**	**1508.58**

Table 1 provides overall summary statistics for the aggregated number of openings, stations, kilometers, and metro areas receiving at least one transit expansion. The pace of overall transit growth has remained relatively constant during the 21st century, but a few trends are evident: LRT was the dominant mode of choice for expansion at the start of the century, but has gradually declined in favor of more BRT expansion. HRT has seen comparatively less expansion compared to LRT and BRT.

Figure 1

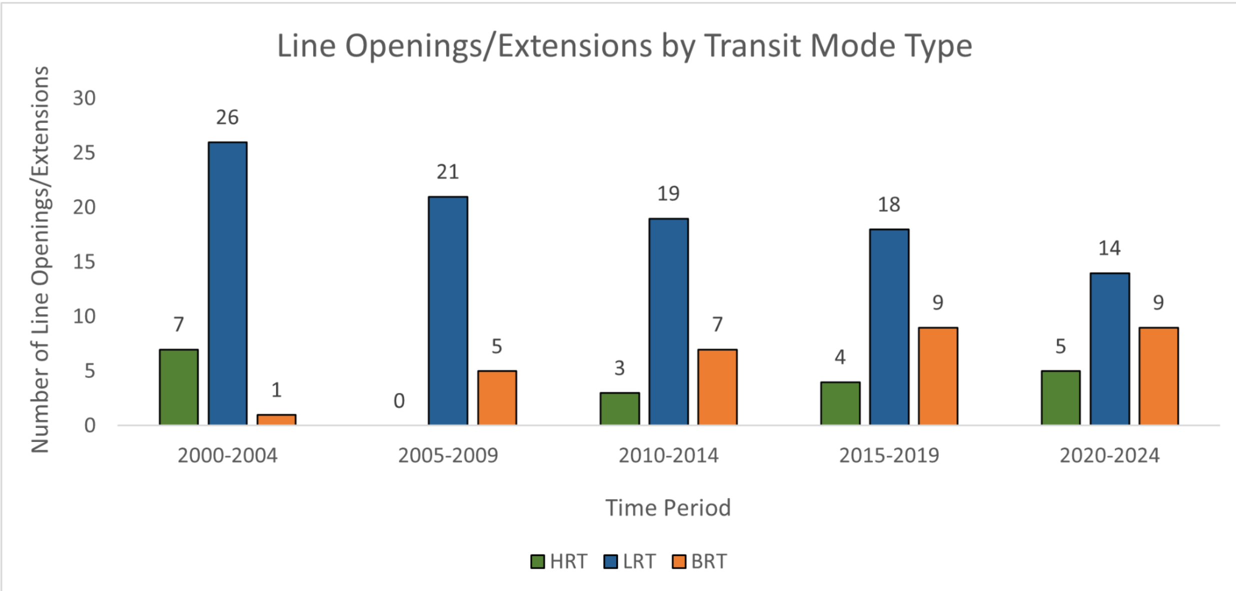


Figure 1 visually represents the trends visible in Table 1 for the number of expansions by mode type. There has been a continuous decrease in the number of LRT openings since the turn of the century, while BRT expansions have grown or stayed the same every five-year period.

Figure 2

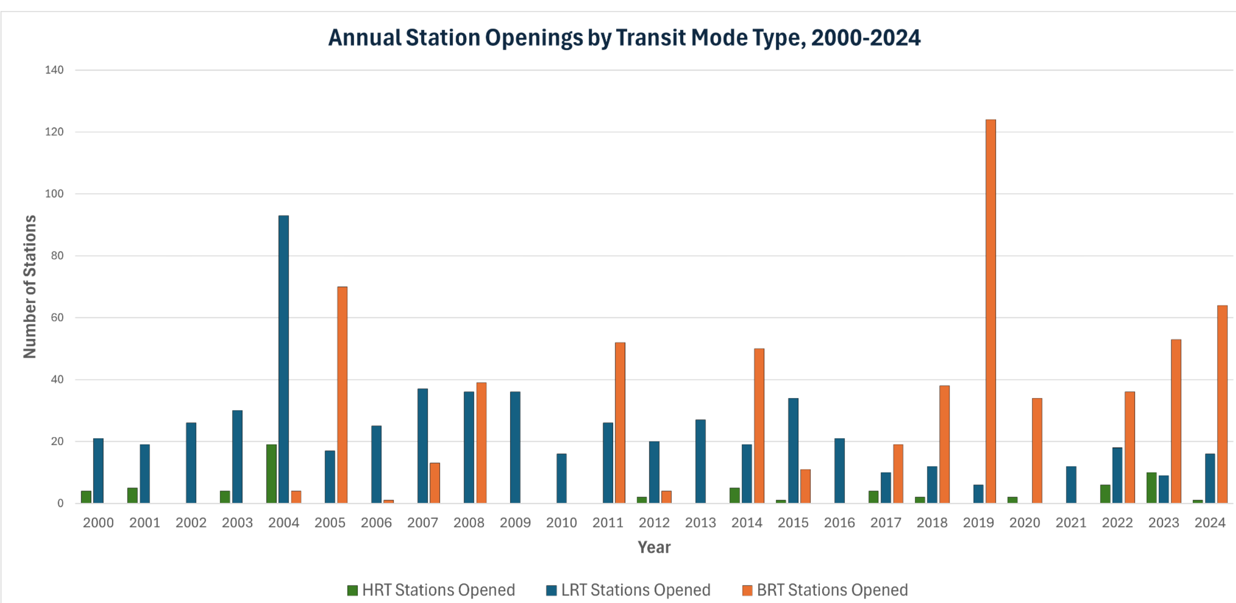


Figure 2 displays a few key trends in transit station openings over the last 25 years: While LRT expansion was the most significant form of transit investment at the start of the 21st century (in terms of the number of stations opened), LRT has gradually tapered off. At the same time, BRT expansion has skyrocketed, particularly over the last 10-15 years. HRT expansion has not seen significant growth during the 21st century.

Figure 3

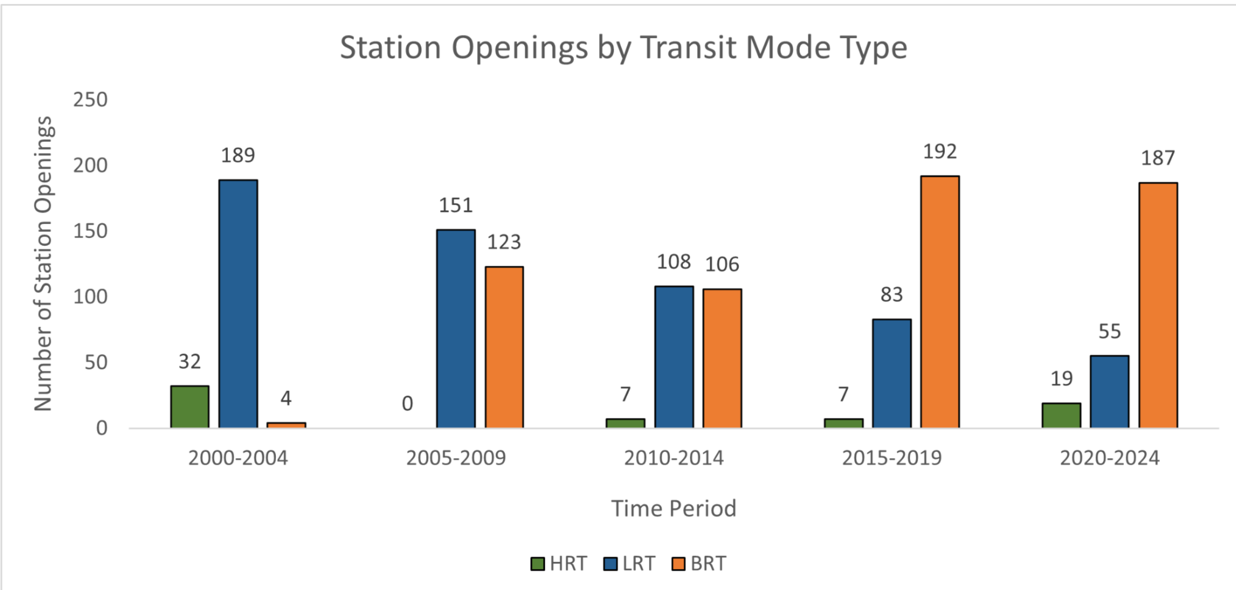
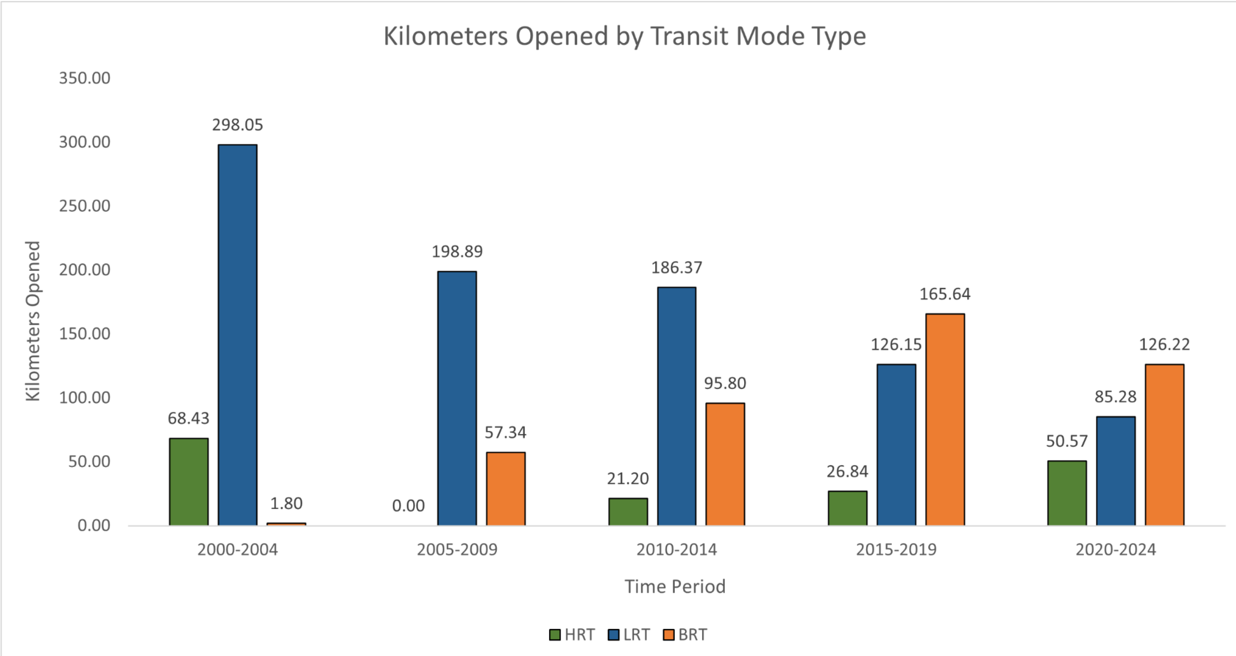


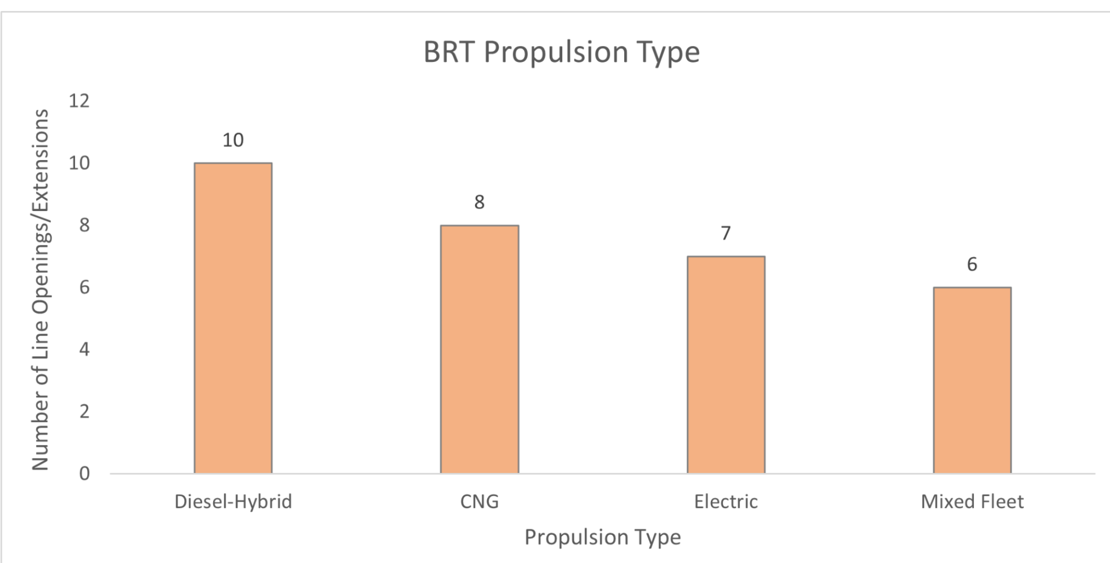
Figure 3 presents station openings broken down by transit mode type in five-year intervals. We see similar trends to the previous graphs: HRT openings are stagnant, LRT construction has fallen in recent years and BRT went from a nascent transit type to having the highest rate of expansion among the three modes.

Figure 4



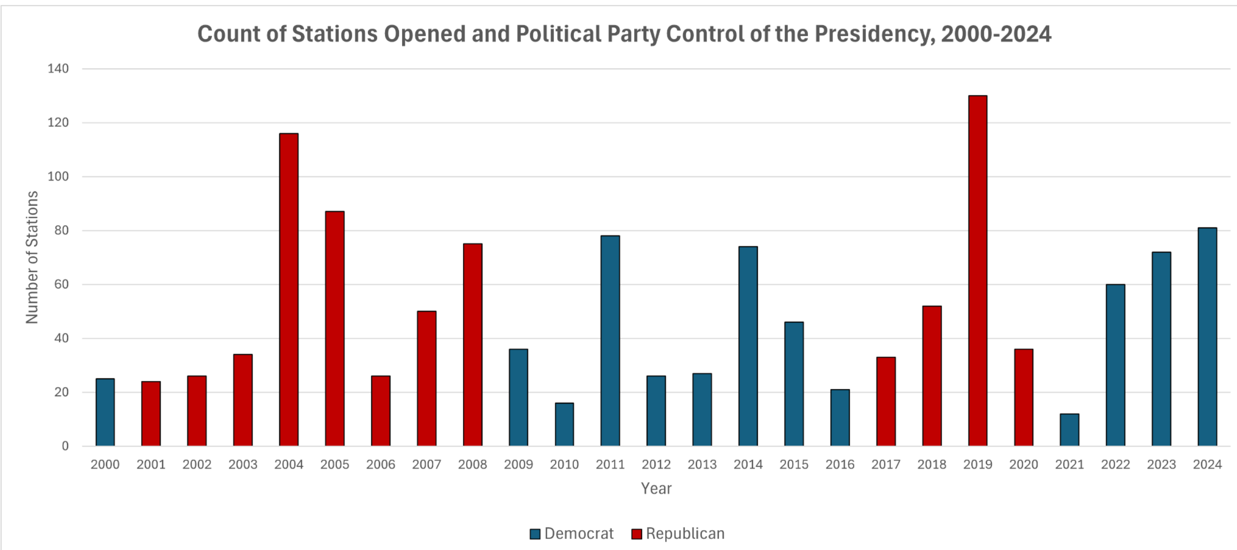
We see similar trends when viewing the total number of kilometers opened, broken down by transit mode. The kilometers of track opened for LRT expansions far outpaced that of both HRT and BRT, but BRT has gradually gained steam especially over the last decade of growth.

Figure 5



The majority of BRT openings/extensions are operating with buses running on fossil fuels. Less than 25% of expansion projects are using a fully electric bus fleet. Three of the mixed-fleet expansions (all on the MBTA Silver Line in Boston) partially run electric buses, with the rest being diesel-hybrid. The other three mixed-fleet expansions are in Kansas City and are operated by diesel-hybrid and CNG buses.

Figure 6



Across all years, a total of 574 stations opened during Democratic control of the presidency and 689 opened when Republicans controlled the White House. Democrats controlled the presidency for 13 out of the 25 years from 2000 to 2024. Figure 6 spurs a hypothesis worthy of future research: Despite Republicans' traditional stance against public transit, we could be seeing more stations open under Republicans because of a significant time gap between when transit projects are approved (presumably more often under Democratic control when transit-friendly policymakers are in charge) and when they begin revenue service, at which point the presidency may have changed hands to the other political party.

Figure 7

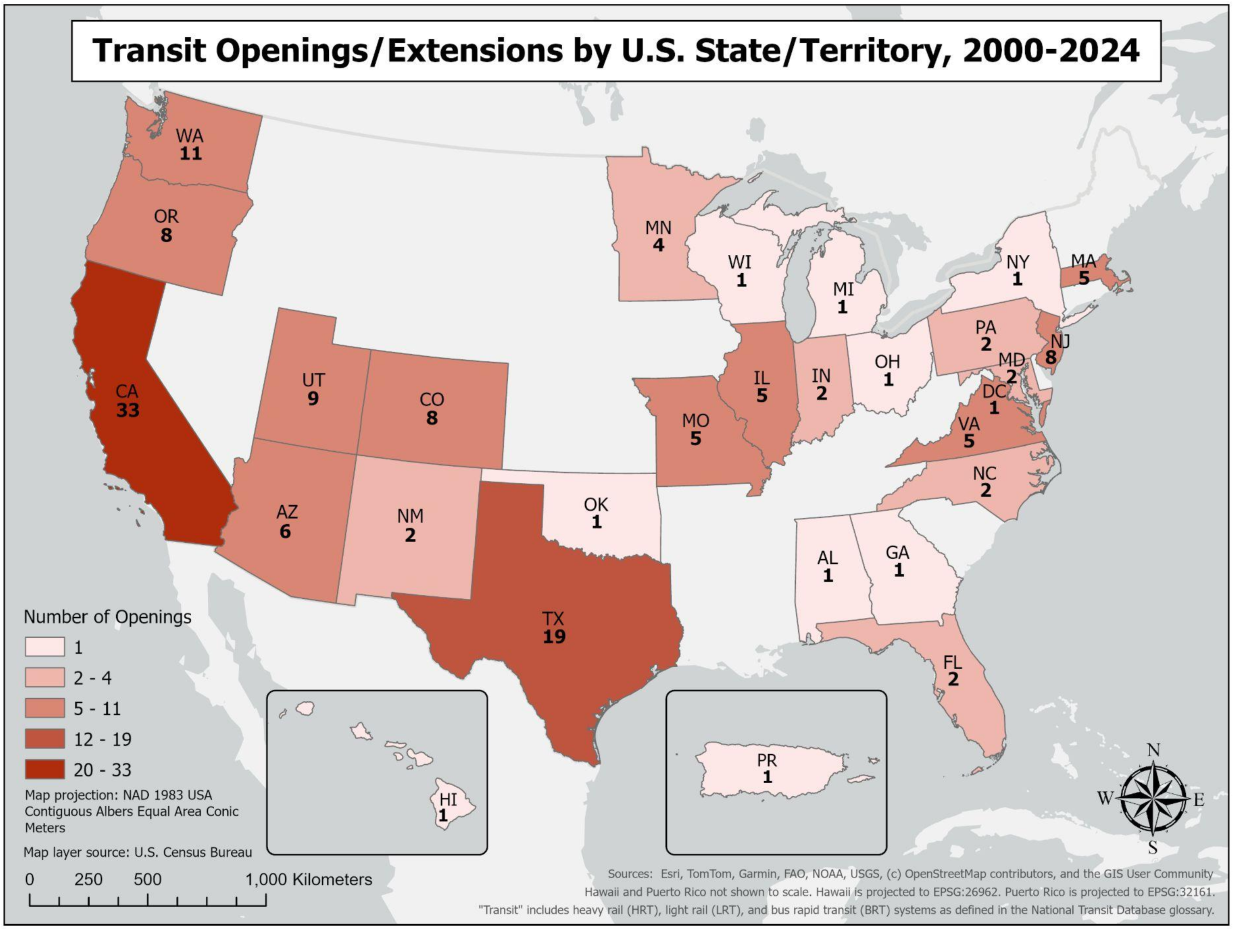
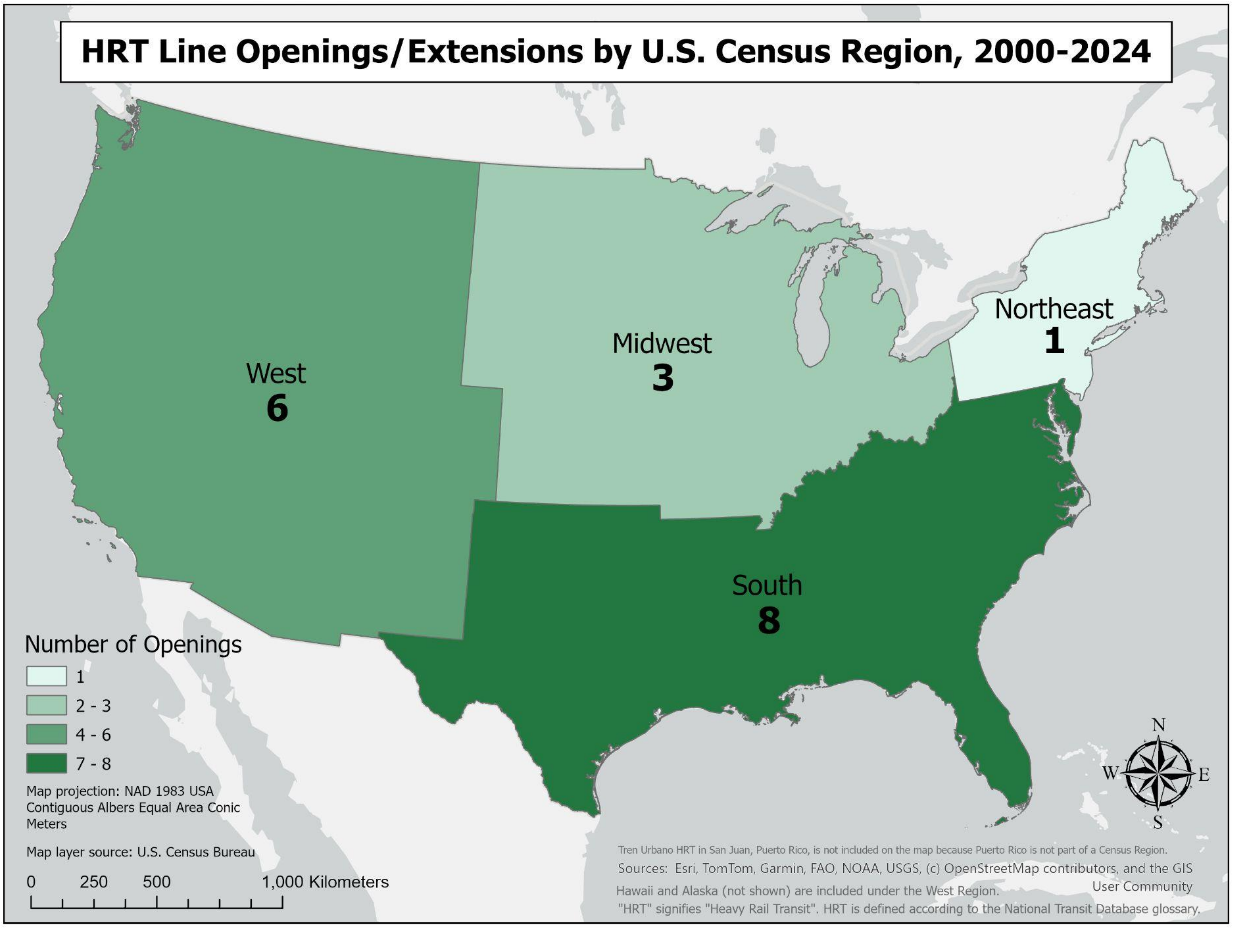


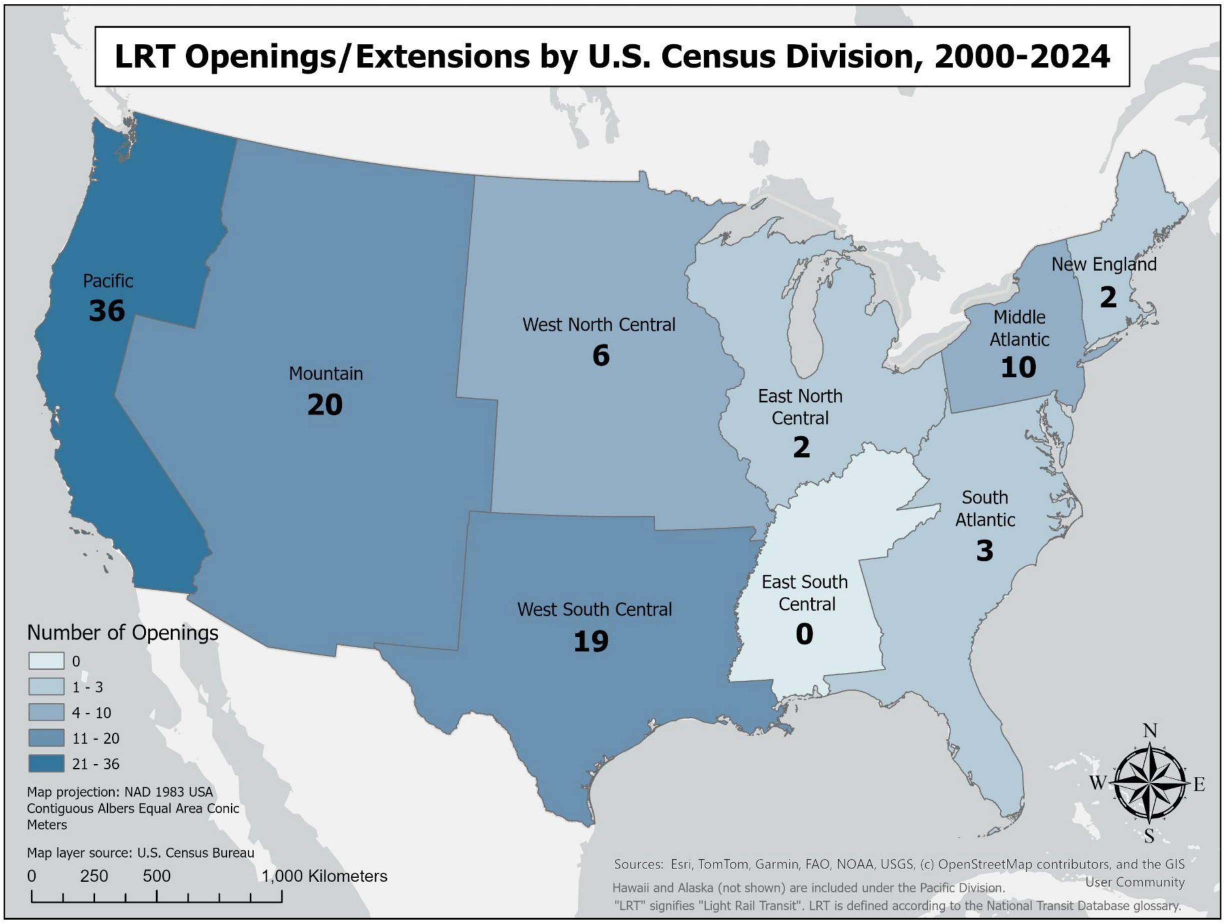
Figure 7 displays geographic patterns of transit expansion over the last 25 years. A disproportionate number of expansions have taken place in the western United States. California leads all states with 33 openings and extensions, followed by Texas with 19. A total of 21 states did not open an HRT, LRT, or BRT project. For more information about how we determined which state to assign a transit expansion that spanned multiple states or opened in the suburbs of a city in another state, refer to Point 2.0 in *Appendix B – Data Practices.*

Figure 8



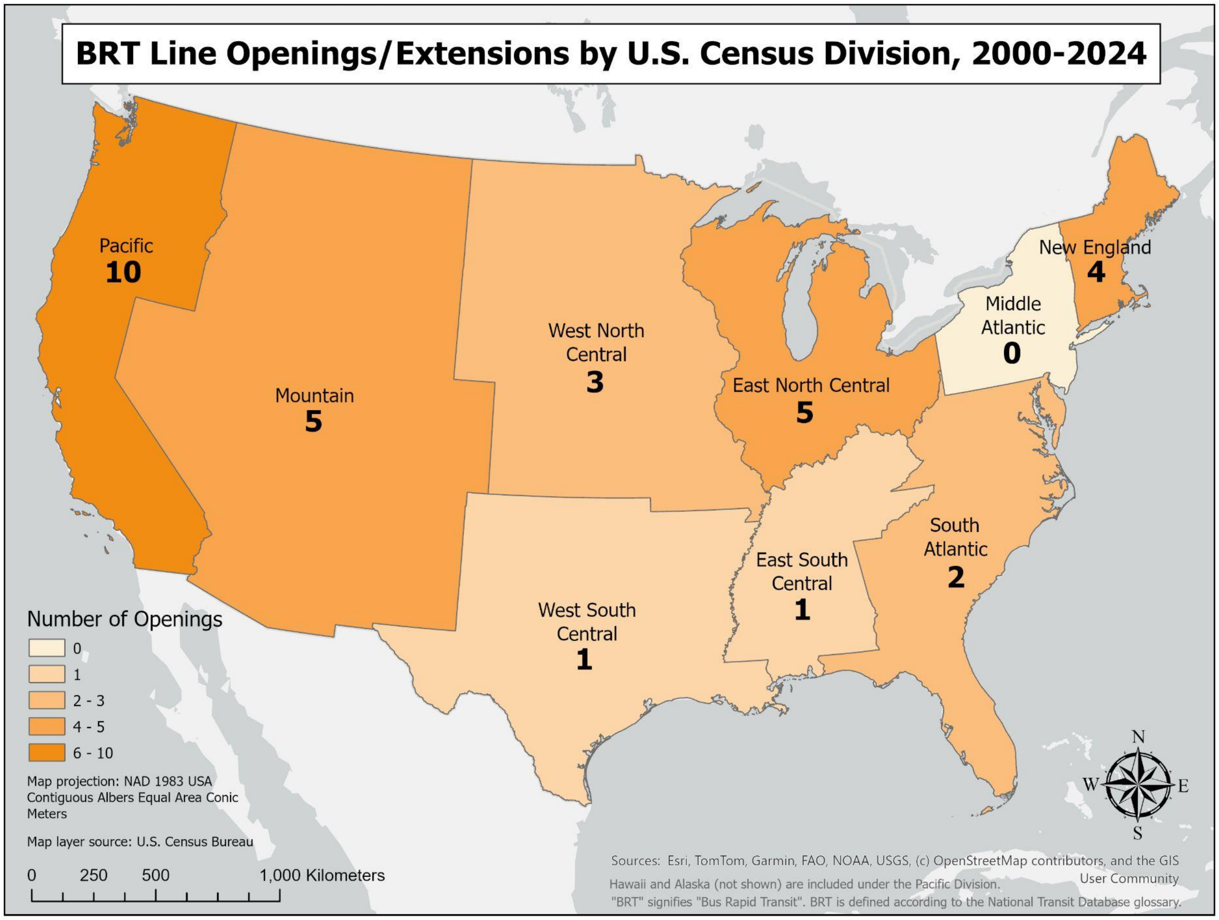
The geographic distribution of heavy rail expansion projects shows the southern and western U.S. comprising the vast majority of HRT growth over the last 25 years. Figure 8 is split by the U.S. census-defined four regions of the country—Northeast, Midwest, South, and West. The Northeast experienced the smallest amount of HRT expansion, with only the MTA's 2nd Ave Subway in New York City appearing on the map. Three infill stations in Chicago compose the only three openings in the Midwest. All of the growth in heavy rail from the South and West has been with the transit systems of Washington, D.C., Miami, Atlanta, Los Angeles, San Francisco, and Honolulu.

Figure 9



While Figure 8 splits the U.S. by census region, Figure 9 breaks the country down further by census division. Given the much larger quantity of LRT openings and extensions compared to HRT growth, the finer detail is beneficial for spatial visualization. Figure 9 shows the geographic bias in LRT expansion to the western United States, with 56 openings and extensions from the Pacific and Mountain divisions, which is more than half of all LRT projects. Texas, in West South Central, is another source of nearly 20 LRT expansions. The Middle Atlantic's relative strength compared to other eastern divisions is powered by the several light rail projects in New Jersey. East South Central—composed of the states of Mississippi, Alabama, Tennessee, and Kentucky—was the only division that did not have at least one LRT opening or extension.

Figure 10



We see the strength of the western United States with BRT projects; nearly half of all BRT growth has occurred in the Pacific and Mountain divisions. But the geographic distribution of BRT expansion is more evenly distributed than both HRT and LRT. Four separate metropolitan areas received new BRT service in the East North Central division. Three of the four BRT expansions in New England occurred in Boston, with the other being in Hartford, Connecticut. The Middle Atlantic, composed of New York, New Jersey, and Pennsylvania, was the only division that did not experience BRT growth.

Figure 11

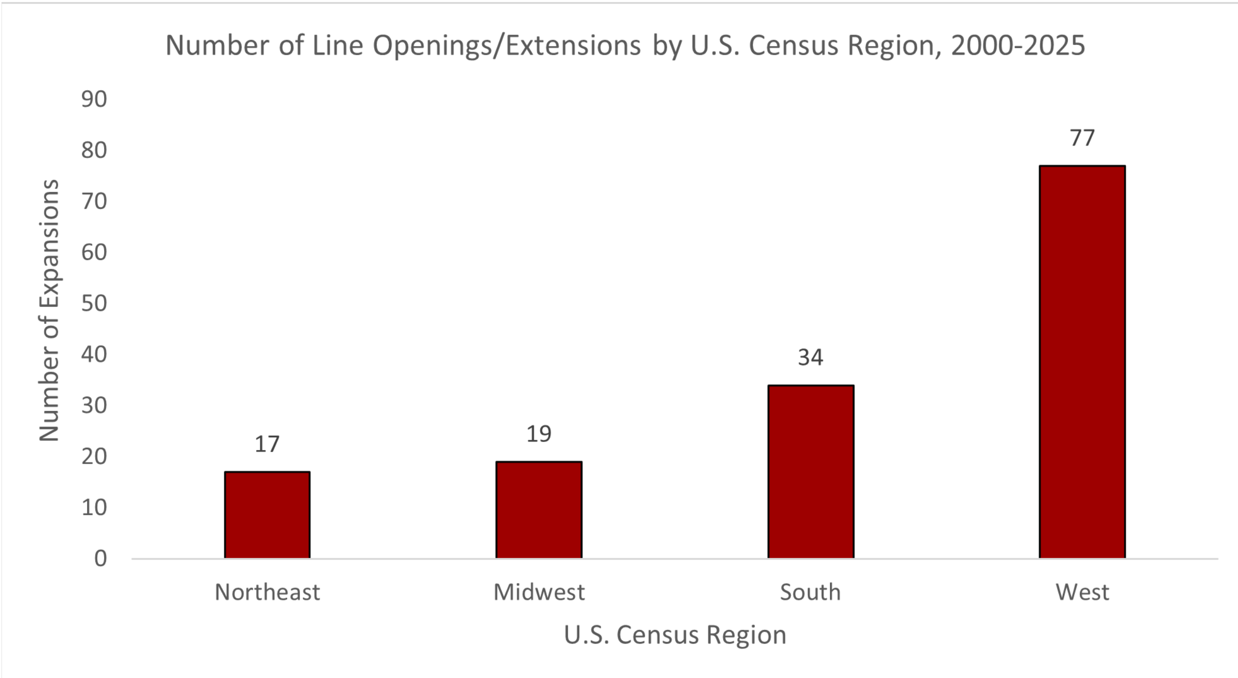
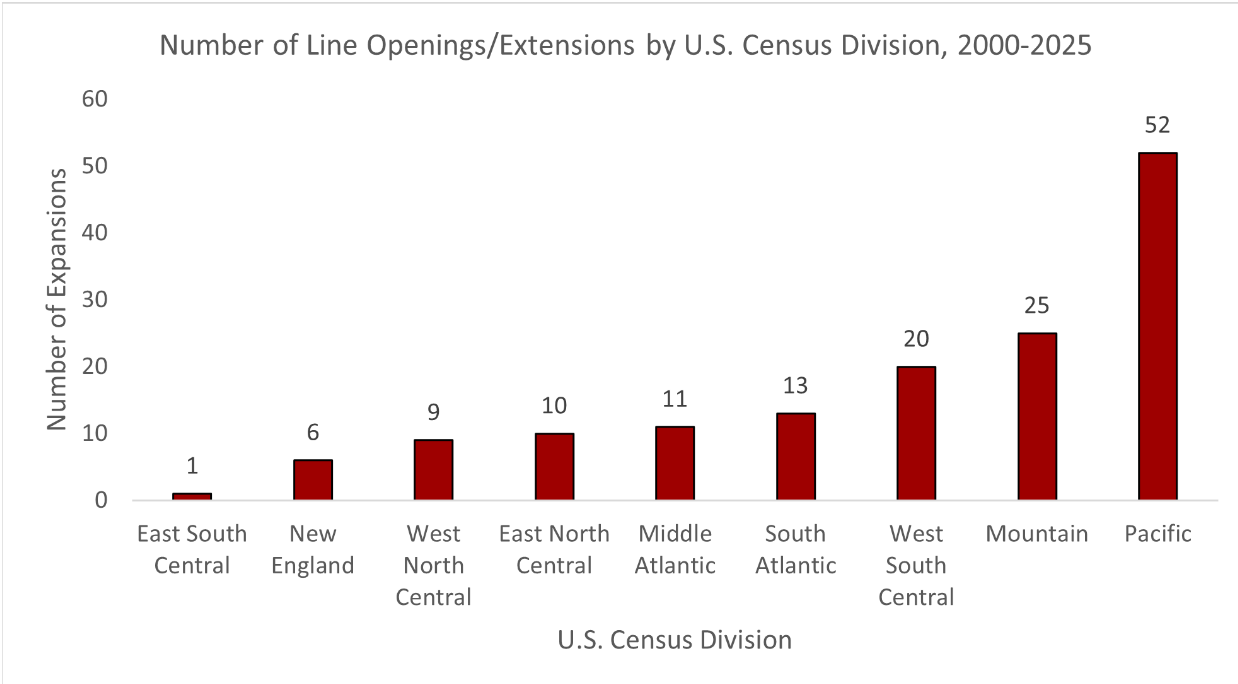


Figure 11 breaks down the total number of line expansions (infill stations are included) by census region. With 77 openings and extensions, the West is home to over half of all HRT, LRT, and BRT projects over the last 25 years. The Northeast, Midwest, and South combine for 70 expansions.

Figure 12



Breaking down the country further by census division, we see the dominance of transit expansion in the western United States. The Pacific and Mountain divisions lead the nation, followed by the southern divisions.

5. Future Research and Conclusion

This project lays the foundations for future research related to the impacts and co-benefits of public transit expansion in the United States. We hope that researchers will be able to use the shapefiles produced here as a core facet of research that analyzes transit's impacts on public health, land use, population and demographic changes, socioeconomic status, displacement, and more. The attribute tables could be expanded (or associate tables created) to include more information from General Transit Feed Specification (GTFS) data. GTFS files typically include scheduled and actual arrival and departure times, complete timetables, fare structures, and more.

It is possible to continue this foundational research in other countries. Transitland has GeoJSON files available in Canada, much of Europe, parts of Asia, Australia, and New Zealand. While many countries classify their transit modes differently, much of the protocol used for generating shapefiles in the United States could be applied to work in other countries. Transitland still lacks information from most countries, however. More research would be required to locate alternative sources for geospatial transit data in other parts of the world. Geospatial data from GTFS—which is more widely available in over 100 countries—could be a viable alternative data source.

For more information on the download, modification, and distribution of elements of this project, refer to *Appendix D – Data User Agreement.*

References

Aftabuzzaman, M., Currie, G., & Sarvi, M. (2010). Evaluating the congestion relief impacts of public transport in monetary terms. Journal of Public Transportation, 13(1), 1-24.

Albalate, D., & Fageda, X. (2019). Congestion, road safety, and the effectiveness of public policies in urban areas. Sustainability, 11(18), 5092.

Beaudoin, J., Farzin, Y. H., & Lawell, C. Y. C. L. (2015). Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality. Research in Transportation Economics, 52, 15-22.

Delmelle, E. C. (2021). Transit-induced gentrification and displacement: The state of the debate. In Advances in transport policy and planning (Vol. 8, pp. 173-190). Academic Press.

FTA. (2025). National transit database (NTD) glossary | FTA. Federal Transit Administration. https://www.transit.dot.gov/ntd/national-transit-database-ntd-glossary

Gendron-Carrier, N., Gonzalez-Navarro, M., Polloni, S., & Turner, M. A. (2022). Subways and urban air pollution. American economic journal: Applied economics, 14(1), 164-196.

Hu, N., Legara, E. F., Lee, K. K., Hung, G. G., & Monterola, C. (2016). Impacts of land use and amenities on public transport use, urban planning and design. Land use policy, 57, 356-367.

Hurst, N. B., & West, S. E. (2014). Public transit and urban redevelopment: The effect of light rail transit on land use in Minneapolis, Minnesota. Regional Science and Urban Economics, 46, 57-72.

Kim, J., Lee, J., Chung, S., & Jang, K. (2022). Effects of built environment in subway stations on pedestrian injuries. Journal of Transport & Health, 26, 101389.

Li, S., Liu, Y., Purevjav, A. O., & Yang, L. (2019). Does subway expansion improve air quality?. Journal of Environmental Economics and Management, 96, 213-235.

Weisbrod, G., & Reno, A. (2009). Economic impact of public transportation investment (pp. 1-7). Washington, DC: American public transportation association.

Xiao, C., Goryakin, Y., & Cecchini, M. (2019). Physical activity levels and new public transit: a systematic review and meta-analysis. American journal of preventive medicine, 56(3), 464-473.

Appendix A – National Transit Database Definitions

The following are the official definitions of heavy rail, light rail, and bus rapid transit from the National Transit Database (NTD) glossary:

- Heavy Rail: “A transit mode that is an electric railway with the capacity for a heavy volume of traffic. It is characterized by:
  - High speed and rapid acceleration passenger rail cars operating singly or in multi-car trains on fixed rails;
  - Separate rights-of-way (ROW) from which all other vehicular and foot traffic are excluded;
  - Sophisticated signaling; and
  - High platform loading” (FTA, 2025).
- Light Rail: “A transit mode that typically is an electric railway with a light volume traffic capacity compared to heavy rail (HR). It is characterized by:
  - Passenger rail cars operating singly (or in short, usually two car, trains) on fixed rails in shared or exclusive right-of-way (ROW);
  - Low or high platform loading; and
  - Vehicle power drawn from an overhead electric line via a trolley or a pantograph” (FTA, 2025).
- Bus Rapid Transit: “Fixed-route bus systems that operate at least 50 percent of the service on fixed guideway. These systems also have defined passenger stations, traffic signal priority or preemption, short headway bidirectional services for a substantial part of weekdays and weekend days; low-floor vehicles or level-platform boarding, and separate

branding of the service. Agencies typically use off-board fare collection as well. This is often a lower-cost alternative to light rail" (FTA, 2025).

Appendix B – Data Practices

For rail openings/extensions:

- 1.1. If a line opening uses existing track at some point along its route, that section of the line is not included in the shapefile. Only the new section of track that opened as part of the line is included. This was done because a lot of transit agencies have changed their line configurations over time, rendering the initial line's route rather meaningless beyond the new section of track. Locating the initial routing on existing track also proved overly complex, requiring extensive scouring of old maps and websites that showed how the line used to function. For the purposes of this research project, it was determined that the inclusion of the existing portion of track that a line used when it first opened was not an integral component to studying the environment and health outcomes associated with the openings/extensions.
- 1.2. For a line extension that connects directly into an existing station (a station that was opened before the extension came online), that existing station is included in the shapefile for the extension. This was done to show how the extension connects with the existing system, without including any additional existing track beyond the existing station.
- 1.3. If a line extension connects to an existing section of track at some distance away from an existing station, the line is cut off in the shapefile at the point that the new section joins the old section. For example, TriMet's MAX Green Line (light rail) splits from the existing track approximately 0.32 miles south of Gateway/NE 99th Transit Center Station—which already existed before the Green Line opened—to continue its journey along the corridor that opened when the line began service. The track is cut off at this separation point in the shapefile, meaning that Gateway Transit Center is not

included as a stop. This was done to maintain consistency with the principle that unless a transit line is built directly off of an existing station, no part of the existing system network should be included in the shapefile. Including the existing station if, and only if, the extension is built directly off of it also allows viewers of the shapefiles to distinguish between extensions that are built directly from a station, and extensions that branch off of an existing transit corridor at some point in their route.

- 1.4. Sometimes two lines that share the same set of new track and branch off from each other using different new track opened on the same day (Dallas, Denver). In this case, the new set of track that both lines share was included on each line's map.

For bus rapid transit openings/extensions:

- 1.5. Some cities opened new sections of BRT that existing bus lines began using, with no additional creation of lines after the BRT section (e.g. San Francisco's Van Ness corridor). In this case, only the section of the bus lines that contains the newly built BRT portion was included in the map.
- 1.6. Some cities constructed a BRT segment for the purpose of introducing a new line or service, but those bus lines also extend to areas that were not upgraded or do not meet the necessary criteria to be considered BRT (e.g. the RapidRide G line in Seattle or the ART line in Albuquerque, which have dedicated lanes and other features for much, but not all, of their routes). In this case, the entire line was mapped, including the portions of the line that are not technically BRT. This decision was made to show how transit service on the whole corridor was impacted by the construction of the BRT system; even if the line extends beyond the BRT infrastructure, there was usually an improvement in service along the entire corridor as a result of the BRT construction.

- 1.7. In Madison, WI, the agency introduced BRT service that turns into two local branches at one end of the line. These branches were not included because they lack BRT infrastructure and receive half the amount of service compared to the BRT corridor.
- 1.8. Every bus line that is classified as BRT in the National Transit Database is mapped, with the exception of New York City's Select Bus Service (SBS). Composed of 16 lines, SBS was not included because, although the lines share some characteristics of BRT, they also lack key features that led to the classification from the Institute for Transportation and Development Policy as "Not BRT". The lines also vary significantly from each other, with different elements of BRT infrastructure and services. The inclusion of these 16 lines would have skewed the summary statistics of BRT systems over a bus network that is generally considered to not be BRT.

For mapping:

- 1.9. All systems, with the exception of HART in Honolulu and Tren Urbano in San Juan, are projected using the USA contiguous Albers Equal Area Conic (EPSG: 102003). HART in Honolulu, HI, is projected using the Hawaii Albers Equal Area Conic (EPSG: 3759). Tren Urbano in San Juan, PR, is projected using the Puerto Rico State Plane (EPSG: 32161).
- 2.0. Some expansions span multiple states or exist in the suburbs of a city in another state. For example, St. Louis, Missouri's MetroLink LRT system has multiple expansions that run entirely in the state of Illinois. When mapping the count of transit openings and expansions by state or territory, we labelled each expansion based on the state in which the majority of the project is located. Both of MetroLink's Illinois expansions are listed as being in Illinois, for instance.

For line and stop tables:

- 2.1. The *stop_count* variable is calculated as the total number of stops along a route that have unique names. Some lines, particularly bus lines, have two stops with identical names in nearly the same location to differentiate the two platforms that compose a transit stop. These stops are counted as just one stop. Any stop in the stop file that does not share the name of another stop is considered an individual stop that is counted singularly.
    - 2.11 An exception to the practice on unique stop names is when there is a single station that is counted twice under the same name, but then has additional text in the station name such as “westbound” or “eastbound”. An example of this is the GRTC Pulse BRT system in Richmond, VA, in which every bus stop (except the terminus stops) includes directional text. There is a “Government Center Westbound Station” and a “Government Center Eastbound Station”. These are the same station, but under the naming conventions listed above, would be counted twice. An exception is made for this naming convention and directional text is ignored, so this station (and all others with directional text) is counted as one station.
- 2.2. The existing station included in a shapefile (per point 1.2) is not included in the stop count in the line table. This is because it is not part of the extension.
- 2.3. Many openings/extensions have two or more lines serving the same segment. In this case, the two lines’ headways are combined and the headway figure from the line table displays the combined frequency of both lines. For example, if two lines share a segment

and run at 20-minute intervals offset by 10 minutes from each other, the headway listed in the table would be 10 minutes.

- 2.31. An exception to this rule is when an opening/extension represents an individual line. For example, Albuquerque's ART BRT system is composed of two lines that overlap each other for much of the route. These two lines opened on the same day and are thus represented in their entirety in separate shapefiles. Each line operates every 24 minutes, with a combined headway of every 12 minutes, but because each line is in a different shapefile the listed headway is 24 minutes.
- 2.32. Another exception to this rule is when an opening/extension has multiple lines that have very different headways operating on it simultaneously. An example of this is Hartford's CTFastrak BRT system, which is used by 10 bus lines at varying frequencies. Because a combined headway would be difficult to calculate and likely would not be representative of the frequency that buses run on the corridor given their diverging headways, the most frequent line's headway is recorded as the headway for the segment. In CTFastrak's case, the most frequent line has a headway of every 7.5 minutes during peak hours, so this figure is recorded.
- 2.33. If an opening/extension has two or more lines serving the same segment for more than half of the route length, the headway is combined. If the majority of the opening/extension is served by only one line, the headway is not combined and the headway of the line that encompasses either the majority or the entire length of the segment is recorded.

    - 2.34. If an opening/extension begins operation with a single line but gains another line at any point after the original opening date, the headway of the original line is recorded and the headway of the additional line is ignored. For example, when DC Metro expanded to Largo in 2004, only the Blue Line ran on the extension. But in 2014, the Silver Line began using the same track as the Blue Line and added more service to Largo. The Silver Line's headway is ignored in the reported figure, however, because the table is designed to be representative of an approximate headway that could have been expected when the extension first opened.
- 2.4. Bus rapid transit lines may utilize stops along the route that previously existed before the BRT opened. Due to the challenge of researching when every bus stop along a line opened (significantly more difficult than finding information on rail stations), the recorded year that each stop opened under *year_open* is the year that the line opened. This may not be accurate to every stop as a result, but it is representative of when the BRT line (which is the unit of analysis for most BRT openings/extensions, and not the BRT segment itself) began using each stop.
- 2.5. There are three line shapefiles with unconventional attribute data appearances: "LA_Metro_Gold_Line_Eastside_Extension_2009_Line.shp", "LA_Metro_Gold_Line_Foothill_Extension_2016_Line.shp", and "LA_Metro_Gold_Line_Opening_2003_Line.shp". These three shapefiles contain multiple rows of data in the attribute table instead of the usual single row. They also have two additional rows, "ID" and "SHAPE_LENG". These shapefiles are not sourced from Transitland and are instead sourced directly from Los Angeles Metro; in raw form, they

are standard shapefiles instead of GeoJSON files. These extensions could not be sourced from Transitland because some openings/extensions on the Gold Line (now a combination of the E and A lines) utilized tracks along its right-of-way that are no longer used. Transitland only has JSON files for the current right-of-way, necessitating a different source for these three expansions. The two additional rows described above cannot be removed, as they contain coordinate information that is required for the shapefile to function. The standard columns found in every other line file are included in each row..

- 2.6. For BRT systems, the line file has an additional column named “fuel_type”. This denotes the fuel type that buses which run on the line utilize. Note that the reported fuel type is the current fuel type of the buses, and not the fuel type of the buses when the line first opened or was extended. This is because it is very difficult to find information about what fuel type older buses that are no longer in service used on a particular line.
- 2.7. Some openings/extensions are singular infill stations (stations that open somewhere along an existing right-of-way). Infill stations only have a stop shapefile, and not a line file, because there is no additional right-of-way that opens along with them. Infill stations’ stop shapefile attribute data include an additional column, “headway”. This is the same column that is found in the line file attribute data for expansions that have line files; because infill stations do not have line information, however, this column is added to the stop file’s attribute data.

Appendix C – Miscellaneous Visuals

Figure 1: Sample of Line File Appearance in ArcGIS Pro

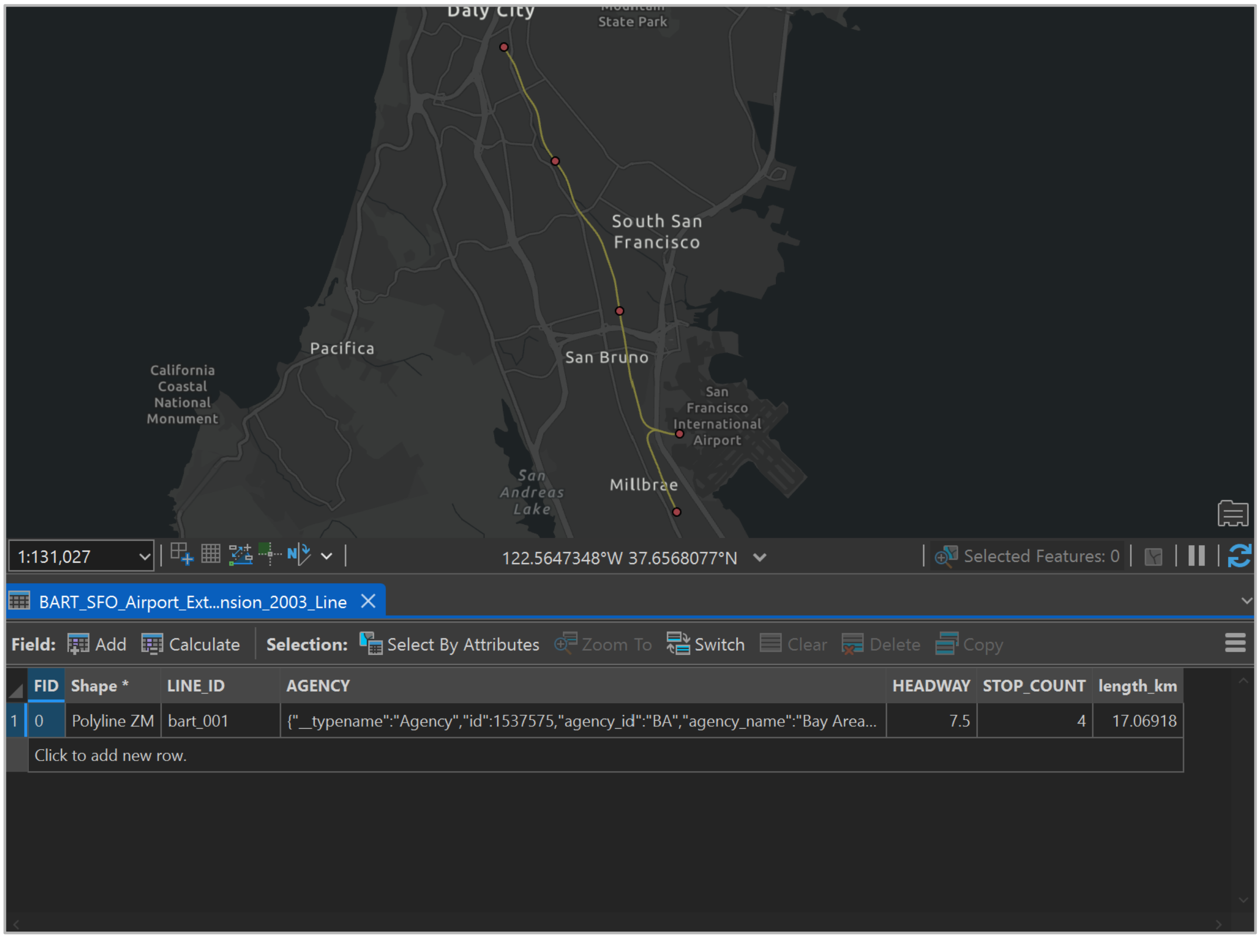
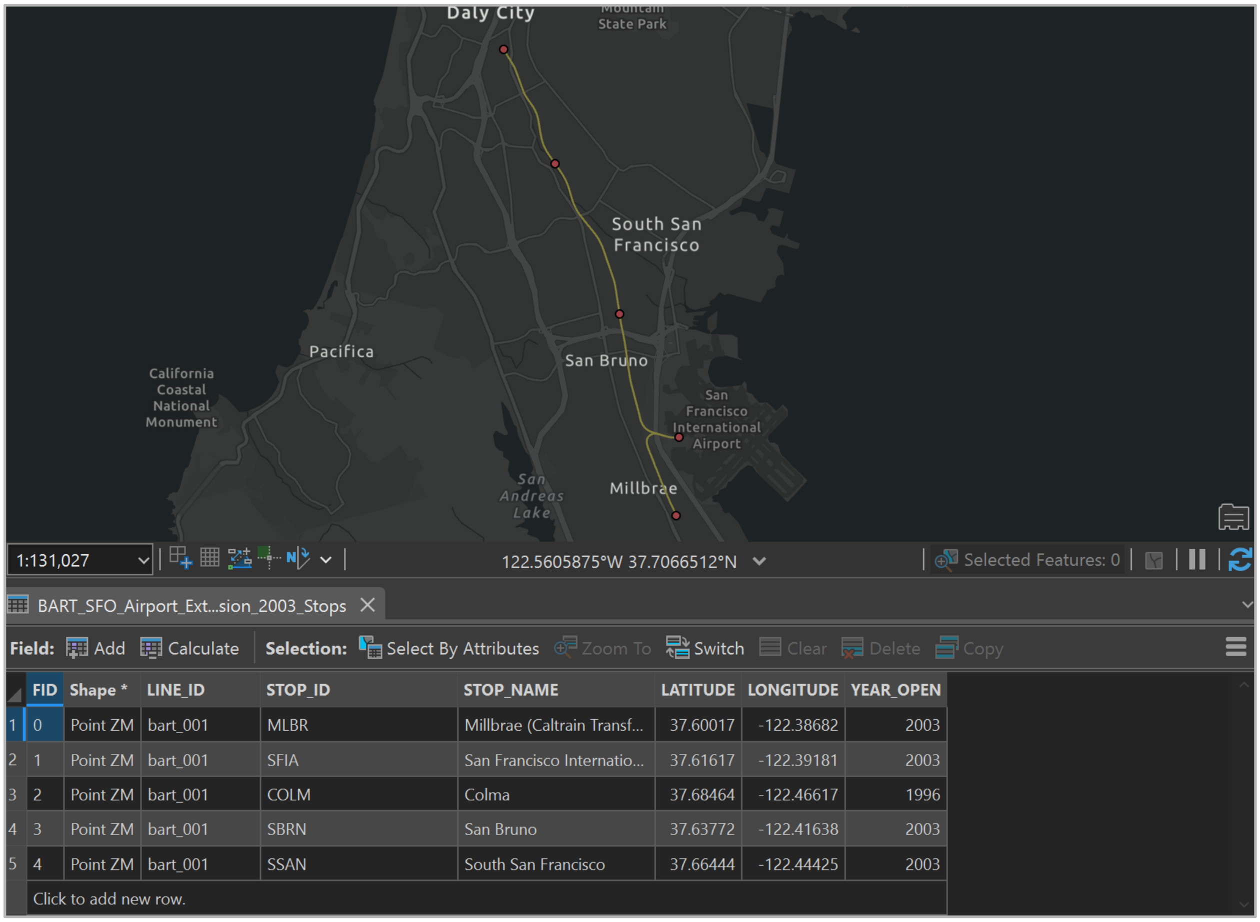


Figure 2: Sample of Stop File Appearance in ArcGIS Pro



	FID	Shape *	LINE_ID	STOP_ID	STOP_NAME	LATITUDE	LONGITUDE	YEAR_OPEN
1	0	Point ZM	bart_001	MLBR	Millbrae (Caltrain Transf...	37.60017	-122.38682	2003
2	1	Point ZM	bart_001	SFIA	San Francisco Internatio...	37.61617	-122.39181	2003
3	2	Point ZM	bart_001	COLM	Colma	37.68464	-122.46617	1996
4	3	Point ZM	bart_001	SBRN	San Bruno	37.63772	-122.41638	2003
5	4	Point ZM	bart_001	SSAN	South San Francisco	37.66444	-122.44425	2003

Appendix D – Data User Agreement

This is an agreement ("Agreement") between you the downloader ("Downloader") and the owner of the materials ("User") governing the use of the materials ("Materials") to be downloaded.

I. Acceptance of this Agreement

By downloading or otherwise accessing the Materials, Downloader represents his/her acceptance of the terms of this Agreement.

II. Modification of this Agreement

Users may modify the terms of this Agreement at any time. However, any modifications to this Agreement will only be effective for downloads subsequent to such modification. No modifications will supersede any previous terms that were in effect at the time of the Downloader's download.

III. Use of the Materials

Use of the Materials include but are not limited to viewing parts or the whole of the content included in the Materials; comparing data or content from the Materials with data or content in other Materials; verifying research results with the content included in the Materials; and extracting and/or appropriating any part of the content included in the Materials for use in other projects, publications, research, or other related work products.

1. Representations

A. In Use of the Materials, Downloader represents that:

1. Downloader is not bound by any pre-existing legal obligations or other applicable laws that prevent Downloader from downloading or using the Materials;
2. Downloader will not use the Materials in any way prohibited by applicable laws;
3. Downloader has no knowledge of and will therefore not be responsible for any restrictions regarding the use of Materials beyond what is described in this Agreement; and
4. Downloader has no knowledge of and will therefore not be responsible for any inaccuracies and any other such problems with regards to the content of the Materials and the accompanying citation information.

B. Restrictions In his/her Use of the Materials, Downloaders cannot:

1. obtain information from the Materials that results in Downloader or any third party(ies) directly or indirectly identifying any research subjects with the aid of other information acquired elsewhere;
2. produce connections or links among the information included in User's datasets (including information in the Materials), or between the information included in User's datasets (including information in the Materials) and other third-party information that could be used to identify any individuals or organizations, not limited to research subjects; and
3. extract information from the Materials that could aid Downloader in gaining knowledge about or obtaining any means of contacting any subjects already known to Downloader.

IV. Representations and Warranties

USER REPRESENTS THAT USER HAS ALL RIGHTS REQUIRED TO MAKE AVAILABLE AND DISTRIBUTE THE MATERIALS. EXCEPT FOR SUCH

REPRESENTATION, THE MATERIALS IS PROVIDED “AS IS” AND “AS AVAILABLE” AND WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, NON-INFRINGEMENT, MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, AND ANY WARRANTIES IMPLIED BY ANY COURSE OF PERFORMANCE OR USAGE OF TRADE, ALL OF WHICH ARE EXPRESSLY DISCLAIMED.

WITHOUT LIMITING THE FOREGOING, USER DOES NOT WARRANT THAT: (A) THE MATERIALS ARE ACCURATE, COMPLETE, RELIABLE OR CORRECT; (B) THE MATERIALS FILES WILL BE SECURE ; (C) THE MATERIALS WILL BE AVAILABLE AT ANY PARTICULAR TIME OR LOCATION; (D) ANY DEFECTS OR ERRORS WILL BE CORRECTED; (E) THE MATERIALS AND ACCOMPANYING FILES ARE FREE OF VIRUSES OR OTHER HARMFUL COMPONENTS; OR (F) THE RESULTS OF USING THE MATERIALS WILL MEET DOWNLOADER’S REQUIREMENTS. DOWNLOADER’S USE OF THE MATERIALS IS SOLELY AT DOWNLOADER’S OWN RISK.

V. Limitation of Liability

IN NO EVENT SHALL USER BE LIABLE UNDER CONTRACT, TORT, STRICT LIABILITY, NEGLIGENCE OR ANY OTHER LEGAL THEORY WITH RESPECT TO THE MATERIALS (I) FOR ANY DIRECT DAMAGES, OR (II) FOR ANY LOST PROFITS OR SPECIAL, INDIRECT, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES OF ANY KIND WHATSOEVER.

VI. Indemnification

Downloader will indemnify and hold User harmless from and against any and all loss, cost, expense, liability, or damage, including, without limitation, all reasonable attorneys’ fees and court costs, arising from the i) Downloader's misuse of the Materials; (ii) Downloader's violation of the terms of this Agreement; or (iii) infringement by Downloader or any third party of any intellectual property or other right of any person or entity contained in the Materials. Such losses, costs, expenses, damages, or liabilities shall include, without limitation, all actual, general, special, and consequential damages.

VII. Dispute Resolution

Downloader and User agree that any cause of action arising out of or related to the download or use of the Materials must commence within one (1) year after the cause of action arose; otherwise, such cause of action is permanently barred.

This Agreement shall be governed by and interpreted in accordance with the laws of the State of California and the United States (excluding the conflict of laws rules thereof). All disputes under this Agreement will be resolved in the applicable state or federal courts of California. Downloader consents to the jurisdiction of such courts and waives any jurisdictional or venue defenses otherwise available.

VIII. Integration and Severability

This Agreement represents the entire agreement between Downloader and User with respect to the downloading and use of the Materials, and supersedes all prior or contemporaneous communications and proposals (whether oral, written or electronic) between Downloader and User with respect to downloading or using the Materials. If any provision of this Agreement is found to be unenforceable or invalid, that provision will be limited or eliminated to the minimum extent necessary so that the Agreement will otherwise remain in full force and effect and enforceable.

IX. Miscellaneous

User may assign, transfer or delegate any of its rights and obligations hereunder without consent. No agency, partnership, joint venture, or employment relationship is created as a result of the Agreement and neither party has any authority of any kind to bind the other in any respect outside of the terms described within this Agreement. In any action or proceeding to enforce rights under the Agreement, the prevailing party will be entitled to recover costs and attorneys' fees.