# WeGo Public Transit

[WeGo Public Transit](https://www.wegotransit.com/) is a public transit system serving the Greater Nashville and Davidson County area. WeGo provides local and regional bus routes, the WeGo Star train service connecting Lebanon to downtown Nashville, along with several other transit services.

The data for this project can be downloaded from [here](https://docs.google.com/spreadsheets/d/1ChPKdwEKmIz\_VUwKIq-qbIVRYkGhxTXy/edit?usp=sharing&ouid=102586165284403190694&rtpof=true&sd=true).

In this project, you'll be analyzing the bus spacing to look for patterns and try to identify correlations to controllable or external factors. Specifically, you'll be using a dataset containing information on the headway, or amount of time between vehicle arrivals at a stop.

There are two main variables you will be studying in this project, headway deviation and adherence.

\*\*Headway\*\* is the amount of time between a bus and the prior bus at the same stop. In the dataset, the amount of headway scheduled is contained in the SCHEDULED\_HDWY column and indicates the difference between the scheduled time for a particular stop and the scheduled time for the previous bus on that same stop.

This dataset contains a column HDWY\_DEV, which shows the amount of deviation from the scheduled headway. \*\*Bunching\*\* occurs when there is shorter headway than scheduled, which would appear as a negative HDWY\_DEV value. \*\*Gapping\*\* is when there is more headway than scheduled and appears as a positive value in the HDWY\_DEV column. Note that you can calculate headway deviation percentage as HDWY\_DEV/SCHEDULED\_HDWY. The generally accepted range of headway deviation is 50% to 150% of the scheduled headway, so if scheduled headway is 10 minutes, a headway deviation of up to 5 minutes would be acceptable (but not ideal).

Another important variable is \*\*adherence\*\*, which compares the actual departure time to the scheduled time and is included in the ADHERENCE column. A negative adherence value means that a bus left a time point late and a positive adherence indicates that the bus left the time point early. Buses with adherence values beyond negative 6 are generally considered late and beyond positive 1 are considered early. However, there is some additional logic where the staff applies waivers to allow early departures, such as an express bus that has already picked up everyone at a park-and-ride lot and is only dropping people off at the remaining stops, and also allows for early timepoint records for all records where TRIP\_EDGE = 2 (end of trip), since it is not a problem if a bus ends its trip early as long as it didn't pass other timepoints early along the way.

* Throughout the dataset
* there was **3.27*%*** *missing adherence information*.
* We removed those **11468** *null values*, instead of converting them **“0”**.
* If we converted them **“0”** it would have *increased* the *on-time* *%*, by **0.8*%***
* Investigate a way to resolve missed actual departure times