

### Why do we need to remove linked trips?

In the original survey data, there were many instances of survey respondents reporting each link of a trip individually. For instance, someone might log a home-to-work trip as 4 separate trips like this:

Table 1: Example Set of Unlinked Trips

Trip Number	Purpose	Mode	Travel Time
1	Go to Work	Walk	5
2	Go to Work	Bus	10
3	Go to Work	Train	35
4	Go to Work	Walk	5

This is actually interesting detailed information to know, but it's not how the survey was designed to be taken. The problem with this kind of trip record is that it's inconsistent with other records in the survey, and different from past travel surveys and other survey methods to which we validate our results. For consistency, we want to identify this trip as a single record, just one home-to-work trip, like below:

Table 2: Linked Version of Table 1 Example

Trip Number	Purpose	Mode	Travel Time	Access Mode	Egress Mode	Transit Line #1	Transit Line #2
1	Go to Work	Train	55	Walk	Walk	KC Metro 40	Link Light Rail

In that record, we still keep track of the access modes, the transit transfers, total travel time, and other fields. The example above is what we call an "unlinked" set of trips. The latest data release replaced these unlinked sets with single "linked" trips.

It was clear in early mode share summaries of the Public Release 1 data that transit and walk mode shares were much higher than expected, based on past surveys and current mode share data like the American Community Survey (ACS). Reflecting on the example unlinked trip set in Table 1, it makes sense that transit and walk shares would be inflated if such trips were present. Mode split for that example is 50% transit and 50% walk, but in reality, this is primarily a transit trip (mode split of 100% transit), where walk-access time is added to the total travel time.

### Identifying unlinked trip sets

Identifying and replacing the unlinked trips sets was challenging and required a combination of automated scripting and manual inspection to catch the many logical exceptions and unexpected trip reporting. Both the scripts and visual inspection process followed the following algorithm for each person's trip records (which is usually a set of several trips in a survey day):

- Compare a person's trip to the subsequent trip
- For these two trips, there is a potential unlinked trip if all the following are true:
  - Either the **mode changes** or the mode is the same transit mode for each trip (e.g., Trip 1 is walk and Trip 2 is bus, *or* Trip 1 is bus *and* Trip 2 is bus),
  - trip **destination purpose is the same** for both trips, and
  - **activity duration** between trips is **under 15 minutes** (meaning that someone walked to the bus station or to their vehicle and at the destination they spent no longer than 15 minutes. This helps avoid linking trips that should actually be considered separate trips.)

This approach helps identify most unlinked trips that were improperly coded by the respondent. Unlinked trips were also flagged any time a respondent marked a purpose of "change modes," which was the intended means of identifying a sub-trip. These were easier to flag since they were self-reported and recorded as expected.

Of course, there were many complications beyond the basic algorithm that required thoughtful exceptions. For instance, a single person is likely to record multiple unlinked trip sets, and it's important to separate them as unique trips rather than linking them all together. This can be done by specifying a minimum activity duration gap between trips, which we set at 15 minutes. This is tricky though, because there might be transfers in unlinked trip sets that take longer than 15 minutes (e.g., waiting for a late bus). However, if the duration is set any higher, we risk joining separate trips together. Other issues include drop-off and pick-up trips. We actually want to keep these trips separated to understand each leg of the trip. We even had to come up with logic to deal with long-distance, out-of-town trips, so we didn't lose the local portion of trips to the airport or the region boundary! While we captured many unlinked trip sets correctly with our scripts, there's still room to improve the logic for similar work on future surveys.

### Creating linked trips from unlinked trips sets

Identifying valid unlinked trip sets was the first step. Joining each together to create single linked trips was the next challenge. Many of these trips were produced automatically with a script, but a large number of the more strange sets had to be produced by hand. The same algorithm was followed either way, as described below:

- Identify the primary trip in the set. This is the trip that will contain most of the data for the new linked trip. Rather simply, the longest distance trip of the set was selected as primary trip record.
- Concatenate the mode strings for the set and add to a new field on the primary trip record. This serves a flag for new linked trips and keeps track of the original set configuration for referencing.
  - e.g., "walk-bus-walk"

- Sum the following fields for all trips in the set and replace the existing values for the primary trip:
  - trip distance
  - trip time
- Replace primary trip start time with start time from first record in set.
- Replace primary trip end time with end time from last record in set.
- Replace origin fields (origin TAZ, city, ZIP, location type, etc.) on primary trip with fields from the first trip in set.
- Replace destination fields (origin TAZ, city, ZIP, location type, etc.) on primary trip with fields from last trip in set.
- Combine all transit line and transit system fields from the set into the primary trip
  - get all unique transit line and system values for all trips in the set, and populate them (in order) in the primary trip's fields.