STRIDE
Southeastern Transportation Research, Innovation, Development and Education Center

Making the most of Automated Passenger Counter (APC) data: A Standardized Approach

Sanskruti Joshi, Simon Berrebi, Kari E. Watkins
Georgia Institute of Technology

Background

1. Transit ridership is decreasing across all modes and bus ridership is at its lowest since 1965.

2. While the bus ridership is going down, there has been an increase in vehicle miles traveled which is alarming for cities as it creates externalities such as traffic congestion, pollution and more traffic fatalities.

Inconsistencies in data

1. APC data sometimes has duplicate trips due to updates in schedules

2. Naming conventions used in APC are sometimes different than that in GTFS (e.g, route numbers are different in APC and GTFS)

3. There are missing trips in APC if every transit vehicle is not equipped with counters or have flaws in hardware of the counters.

4. Availability of data for same time period for GTFS and APC

Objective

1. Understand APC and GTFS datasets to use them for research

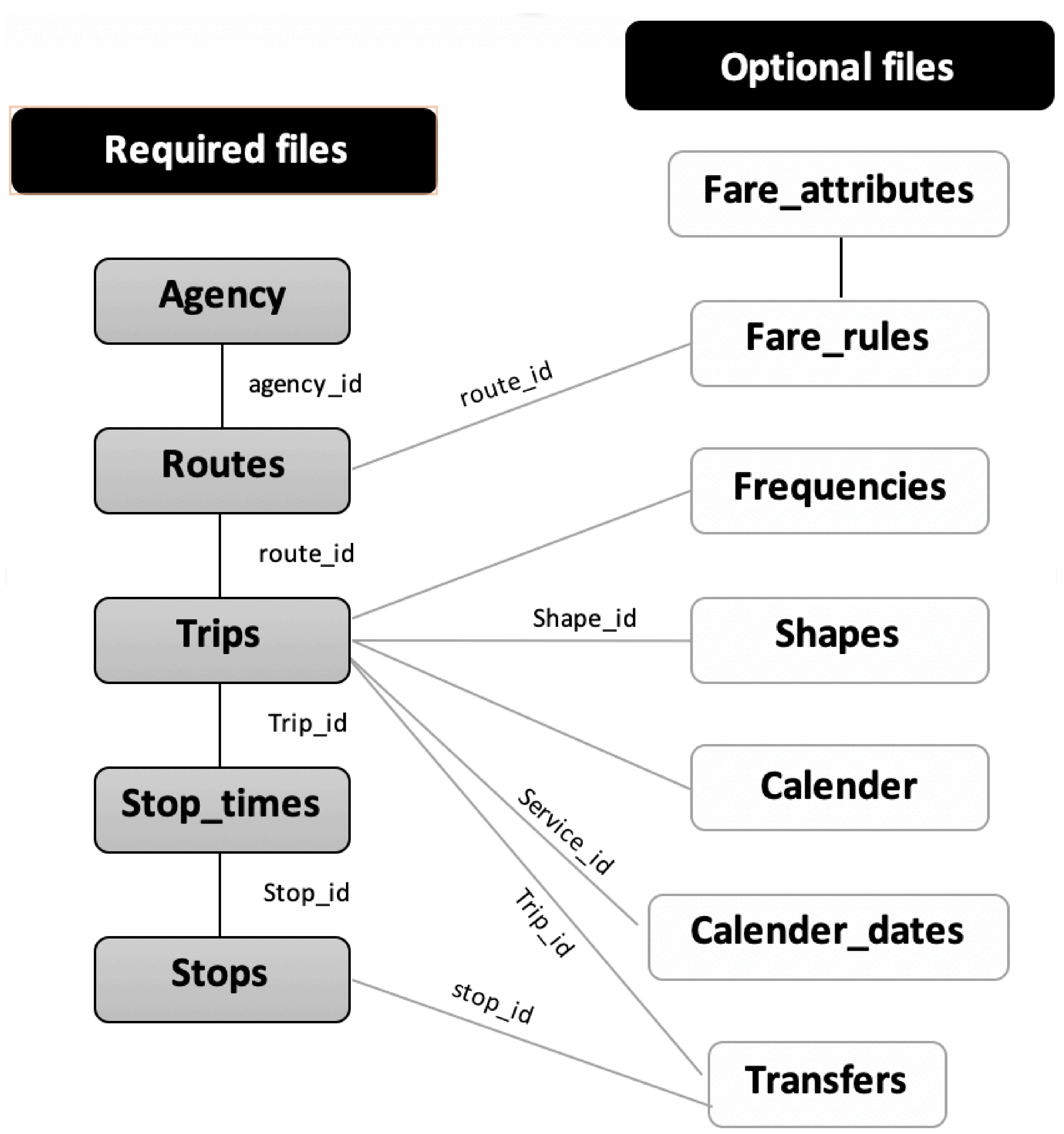
2. Develop a standardized method to process APC and GTFS data that can be used by transit agencies to better manage their datasets and make informed decisions

What are APC and GTFS?

1. Automated Passenger Counters are he hardware to keep track of ridership change on buses or trains and are majorly more accurate than manual ride checks.

2. General Transit Feed Specification (GTFS) which is a platform and a standard format where transit agencies can publish their information about their schedules, stops, number of trips, fares and location of stops.

Components of GTFS



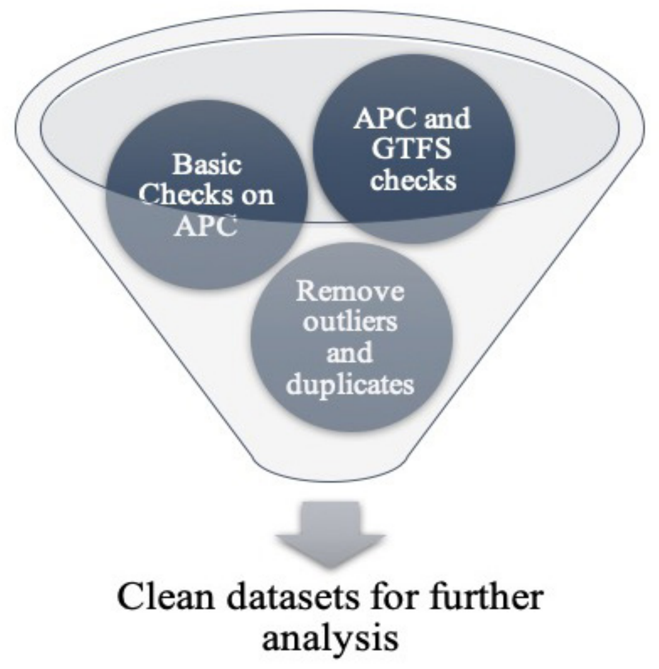
Connections between APC and GTFS

1. APC data generally has stop, route, direction, markup and trip information that can be directly matched with information from GTFS files for the same markup.

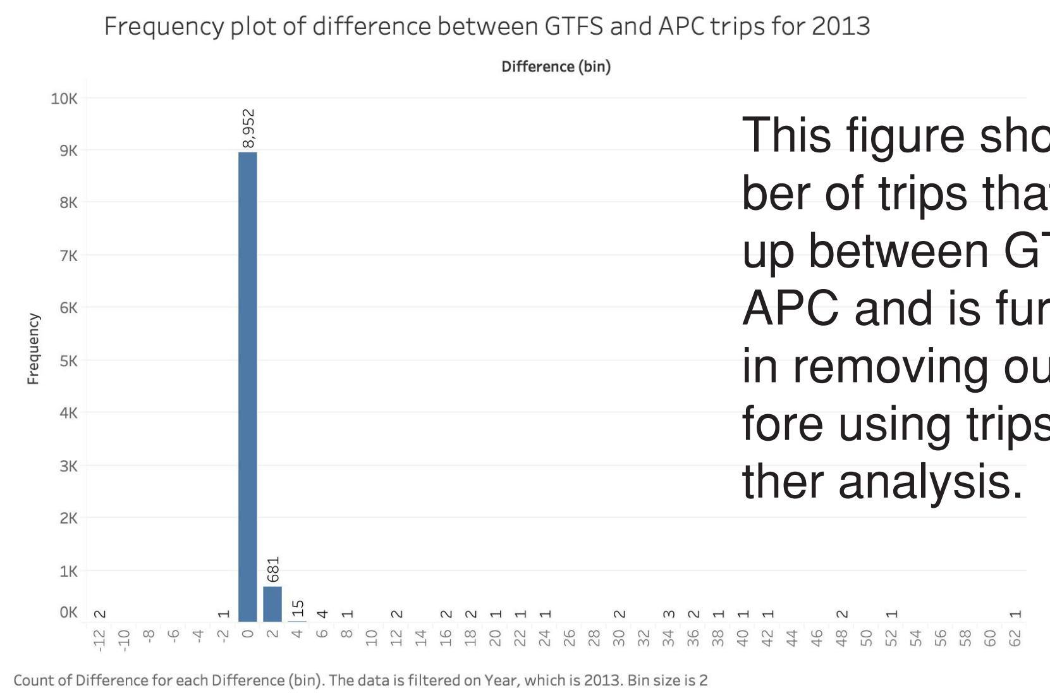
2. For spatial analysis, stop_ids in APC can be matched with latitude and longitude in stops file of GTFS

3. APC has information related to boardings and alightings that can be matched up with stop, route and direction information from GTFS for statistical as well spatial analysis.

Methodology for data cleaning and checking


Basic Checks on APC
APC and GTFS checks
Remove outliers and duplicates
Clean datasets for further analysis

Frequency plot of difference between GTFS and APC trips for 2013



Count of Difference for each Difference (bin). The data is filtered on Year, which is 2013. Bin size is 2

This figure shows number of trips that match up between GTFS and APC and is further used in removing outliers before using trips for further analysis.

Basic checks: Checks on APC data to understand the distribution of raw data which is useful to keep a check on obvious errors.

APC and GTFS checks: Joining APC with GTFS helps in finding missing data from APC compared to GTFS.

Removing outliers such as stop-route-direction combinations that have very high difference in number of trips in GTFS and APC.

Duplicates in trips in APC could be due to change in schedule that is accounted for twice in APC data which is removed by taking into account arrival times of trips for the same stop-route-direction combinations