Transit data

Pramesh Kumar

IIT Delhi

July 23, 2024

Need for data

For planning an efficient, effective, reliable, and productive transit service, agencies require accurate data about operations and usage of PT.

Data collected by the agencies

1. Supply
   - Stop/station facility (location, capacity, dimensions, amenities, etc.)
   - Routes (length, geometry, schedule, type of service, ROW, speed limit, signal operations, etc.)
   - Vehicle data (dimensions, design and performance, age, condition, capacity, etc.)
   - Fare (type, collection method, etc.)
   - Speed, travel time and delay
   - Others
2. Demand
   - Passenger boarding and alighting counts at stops
   - Passenger load on various sections of a route
   - Passenger transfer counts
   - Passenger activity purpose
   - Passenger Fare usage pattern
   - Passenger trajectories (origin, destination, boarding, alighting, transfer, waiting, in-vehicle, and walking time, etc.)
   - Modal split
   - Temporal and seasonal data
   - Other behavioral data

Data collection methods

- Manual
- Automated

Data collection methods

- Manual
- Automated

Scale

1. System-level
   - required for long-term planning
   - gives an overall service picture
2. Route-level
   - characteristics of specific routes
   - required for long-term planning and maximizing route performance
3. Trip-level
   - trip characteristics
   - required for improving operations

Manual data collection

Point check

- Checker is stationed at a transit stop to perform counts and measurements
- Usually done at stop with highest average load (or multiple stops with heavy passenger loads or transfer points)
- Data collected about load count, arrival and departure time, etc.
- Useful for updating the vehicle size, frequency, and departure time.

Ride check

- Checker rides the transit vehicle to perform counts and measurements.
- At each stop, the checker collects data about boarding and alighting counts, arrival and departure times, fare category, gender, etc.
- Useful for updating the vehicle size, frequency, departure, layover, and running times.

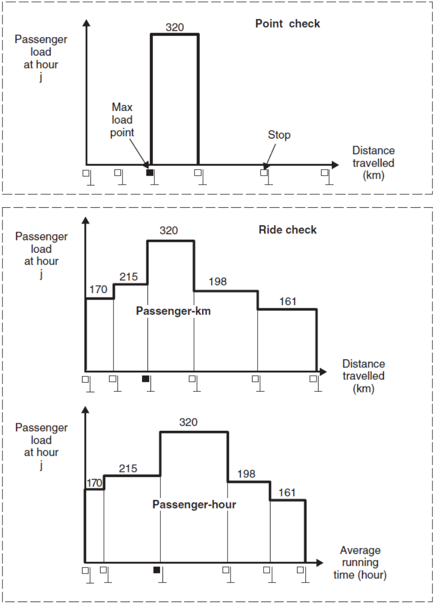


Figure: Point and ride check[1]

[1]Source: Ceder (2019)

Passenger and population surveys

- They can be performed at stops, on-board, terminal or by sending postage-free forms.
- On-board surveys are most common. They can be paper-based or tablet-based.
- Passenger are asked about their socio-economic and trip details, including
  - origin and destination
  - boarding and alighting stops
  - route details
  - transfer information
  - travel time (in-vehicle, waiting, and walking times)
  - purpose of travel
  - fare
- Their attitude and opinions towards new service, fare change, new willingness to pay, etc. can also be collected.
- This OD matrices obtained from surveys are useful for
  - planning new service
  - assess the impact of changes to the existing service (changing route structure, frequency, etc.)

Selecting sample size

Problem statement

We wish to estimate the average load of a transit route in one direction during peak hour so that the error in the estimation is less than or equal to 8 passengers with a confidence level of 95%. How many samples should we collect?

Background

- Assume $X_1, \cdots, X_n$ are sample of size n which are independent and identically distributed (sampling with replacement) random variables.
- Assume that they are normally distributed random variables with $\mathbb{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2$.
- The sample average $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a maximum likelihood estimate of the average value (in our case we want to know the average load).
- $\mathbb{E}(\bar{X}) = \frac{n\mu}{\mu} = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$
- If $Y \sim \mathcal{N}(\mathtt{mean}, \mathtt{var})$ then $\frac{Y - \mathtt{mean}}{\mathtt{var}} \sim \mathcal{N}(0,1)$ (also called standard normal random variable Z).
- $$\mathbb{P}(-2 \leq Z \leq 2) \approx 0.95 \qquad (1)$$

Selecting sample size

We want $\mathbb{P}(|\bar{X} - \mu_{\textsf{load}}| \leq 8) = 0.95 \implies \mathbb{P}(-\frac{8}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{X}-\mu_{\textsf{load}}}{\frac{\sigma}{\sqrt{n}}} \leq \frac{8}{\frac{\sigma}{\sqrt{n}}}) = 0.95 \implies \mathbb{P}(-\frac{8}{\frac{\sigma}{\sqrt{n}}} \leq Z \leq \frac{8}{\frac{\sigma}{\sqrt{n}}}) = 0.95$.

Compare this expression with (1). We have, $\frac{8}{\frac{\sigma}{\sqrt{n}}} = 2 \implies \boxed{n = \frac{\sigma^2}{16}}$.

We may not know the population variance, so we either use sample variance or range ($\approx 4\sigma^2$). Assuming range is 50, then $n = \frac{(\frac{50}{4})^2}{16}) \approx 10$ trips.

Automated data collection

Introduction

- Public transport services have historically been planned with limited knowledge of their customers' travel behavior (using farebox data or surveys).
- Various limitations associated with surveys:
  - Expensive
  - Small sample size
  - Bias
  - General reporting errors
- Automated data can overcome these limitations!
- They indirectly provide a rich source of information about passengers travel pattern on an individual basis.

Transit Automated Data
Automatic Fare Collection (AFC) Data

Metro Transit go to pass

Delhi Smart Card

London Oyster Card

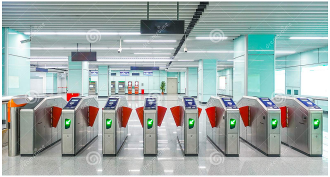
DC Smart Trip Card

Source: https://www.metrotransit.org/passes
https://www.visitlondon.com/traveller-information/getting-around-london/oyster
https://www.wmata.com/fares/smartrip/
https://images.hindustantimes.com

Automatic Fare Collection (AFC) Data

Contactless smart card primarily used for revenue management

- Serial ID assigned to the pass
- Date and time of the tag
- Route information
- Geographical coordinates of the tag
- Open versus closed transit systems
- Traditionally not available in real-time



Source:https://tinyurl.com/2x7mth39

Bus ticketing

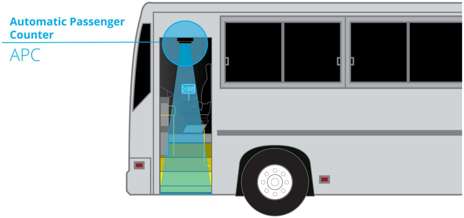
Electronic ticketing in Chennai

Haryana-Chandigarh paper tickets

Automatic Passenger Count (APC) Data

Sensors installed in vehicles with channelized passenger movements. Primarily used for evaluating ridership

- Date and time of operation
- Route, trip and stop information
- Geographical coordinates
- Number of boarding and alighting at every stop
- Passenger load on trains/buses
- Traditionally not available in real-time

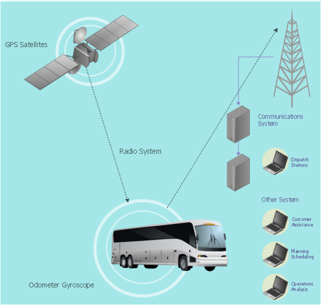


Source: https://www.tsomobile.com/wp-content/uploads/2018/

Automatic Vehicle Location (AVL) Data

GPS-based systems primarily used to provide real-time bus arrival information to passengers.

- Date and time of operation
- Route, trip and stop information
- Geographical coordinates
- Departure and arrival time at bus stops
- Available in real-time



Source: https://conceptdraw.com/a2516c3

General Transit Feed Specification (GTFS) Data

Transit schedule data provided by many transit agencies all over the world. Used by Google Maps to provide directions.

- Agency
- Stops
- Routes
- Trips
- Stop times
- Calendar

Source:https://addtransit.com/blog/2015/

Cellphone applications

Based on cellphone GPS, they track passenger trajectories.

Issues with automated data

- Automated data collection system (ADCS) are designed for specific purpose (e.g., revenue management, online information, etc.)
  - Travel behavior of passengers is not directly observed.
  - Inference methods are required
- Most ADCS are implemented independently
  - Not easy to integrated data
  - Requires new expertise and resources

Potential of ADCS

1. Strategic-level planning: long-term planning
   - Demand aggregation (stop, route, and network)
   - Passenger classification
   - Passenger stop, route, transfer choice
2. Tactical-level
   - Network-level adjustments
   - Assessing reliability
   - Frequency/Schedule changes
   - Where to provide transfer waiting facilities?
   - Planning for special events, bad weather
3. Operational-level
   - Ridership statistics
   - Transfer synchronization
   - Level of service (wait, walk, travel time)
   - Real-time announcements
   - Operations management

Applications: Passenger load

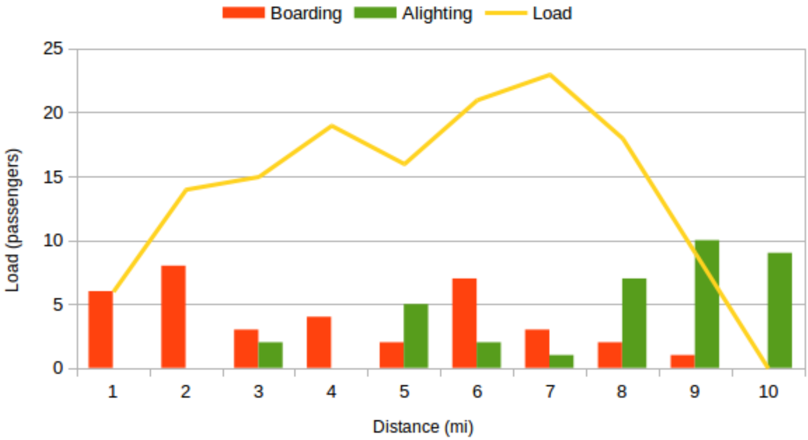


Figure: Sample load profile[2]

[2]Taken from CEGE3201 taught by Prof. Alireza Khani

Applications: OD Route-level

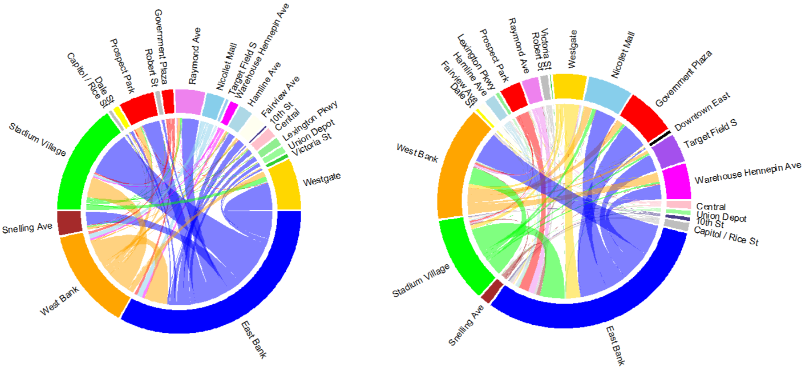


Figure: Passenger origin-destination flow on Metro Green Line light rail during evening peak (EB and WB)

Applications: OD Network-level

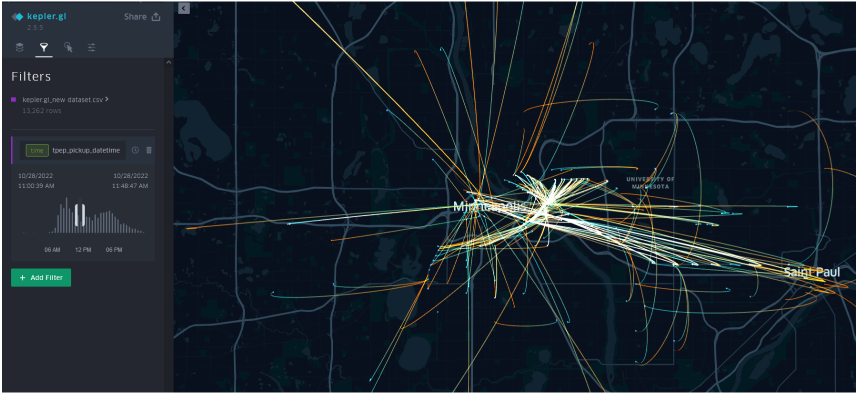


Figure: Time-dependent visualization of passenger trips

Applications: Transfer reliability

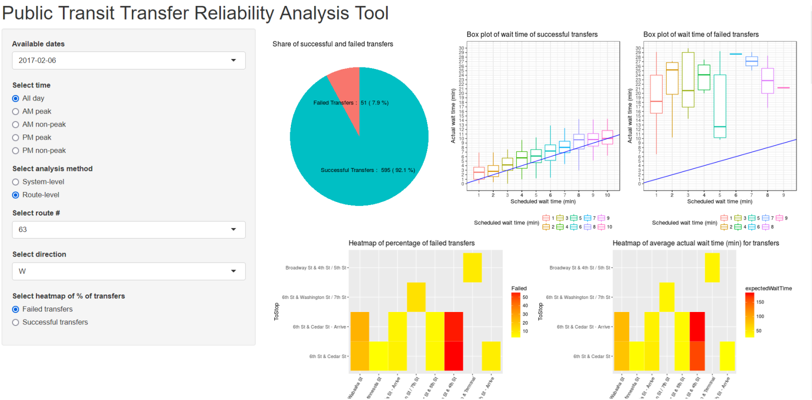
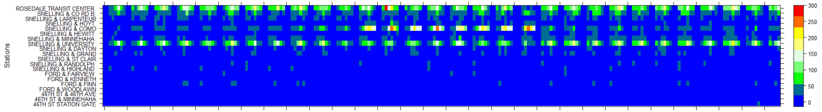
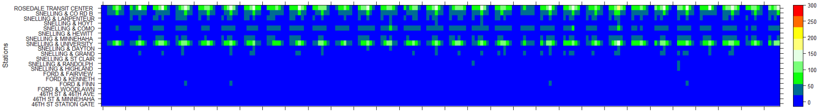


Figure: Transfer reliability

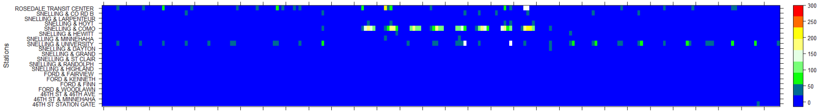
Applications: Evaluating demand for special events



(a) matrix M



(b) matrix L



(c) matrix S

Applications: Transit assignment

Predicts passenger path choice given the network and passenger demand

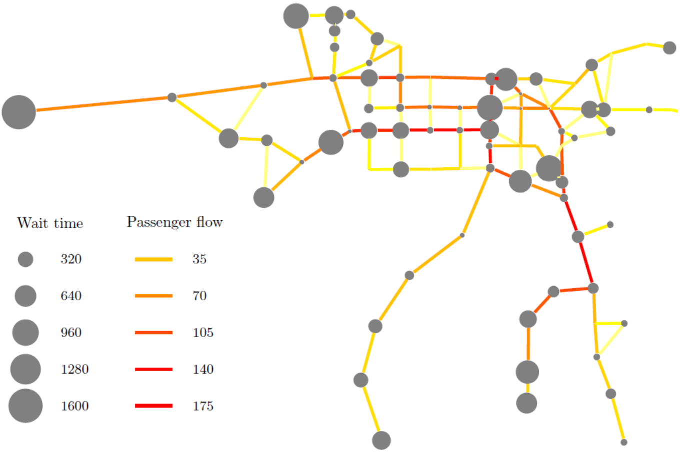


Figure: Transit assignment of Rivera city in Uruguay

Suggested reading

- Vuchic Chapter 1
- Ceder Chapter 2

Thank you!