

Using Google's Aggregated and Anonymized Trip Data to Estimate Dynamic Origin-Destination Matrices for San Francisco

TRB Applications Conference 2017

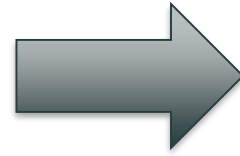
Bhargava Sana, Joe Castiglione,
Dan Tischler, Drew Cooper



SAN FRANCISCO COUNTY TRANSPORTATION AUTHORITY

May 16, 2017

OD Data Collection



- ▶ **Conventional methods**
 - ▶ License plate surveys
 - ▶ Roadside interviews

- ▶ **Emerging passive data collection methods**
 - ▶ Bluetooth detectors
 - ▶ Cell phone Call Detail Records (CDR) data
 - ▶ GPS, Wi-Fi detectors

Google's Better Cities Program

Aggregated and Anonymized Trip (AAT) Data



- ▶ Minimize congestion, improve safety and reduce infrastructure spending
- ▶ Aggregated and Anonymized Trip (AAT) information from location reports
 - ▶ Extract data from moving users
 - ▶ Clean data and snap to road network
 - ▶ Aggregate OD trip counts
 - ▶ Apply differential privacy filters and minimum trips threshold



Table Details: 2015_spring_autumn_copy

Row	origin	destination	via	weight	travel_time_min	start_interval_s	end_interval_s
3279139	70	80		4.120269197703011E-4	null	1428967200	1429008000
3279140	70	80	250_north_of_untermyer_SB	3.900880992332721E-4	null	1428967200	1429008000
3279141	70	79		2.8886334621347487E-4	null	1428967200	1429008000
3279142	70	33		2.8489170848844104E-4	null	1428967200	1429008000
3279143	79	34		2.96296601919155E-4	null	1427793600	1427832000

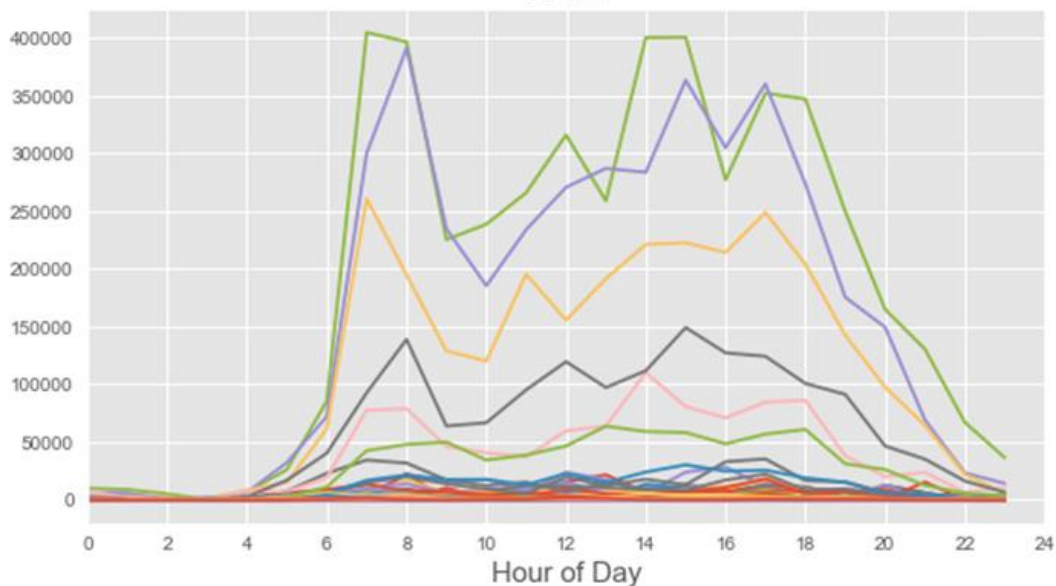
AAT DATA

Google AAT Dataset

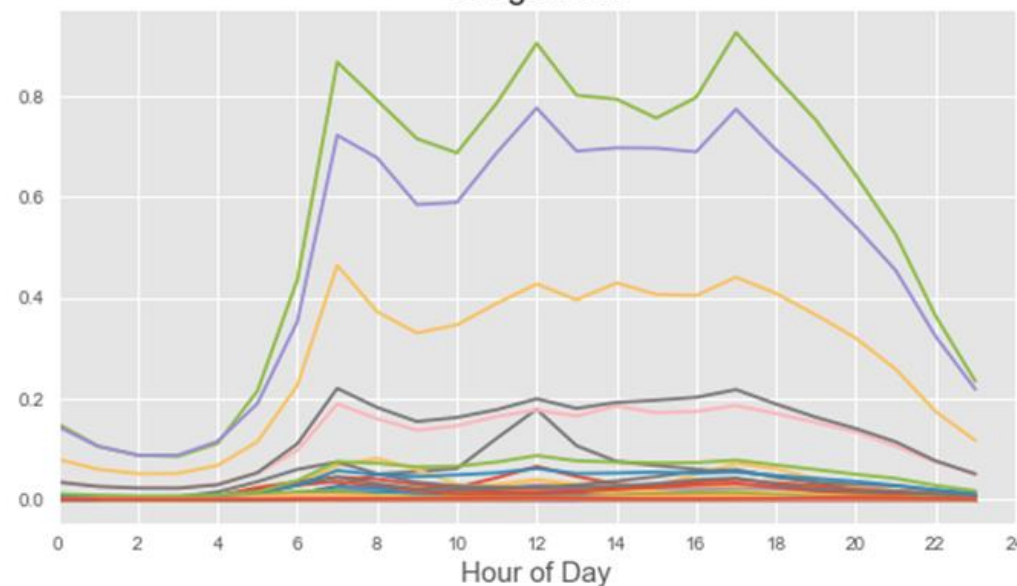


- ▶ Hourly AAT data for six months (Apr-Jun and Sep-Nov 2015)
- ▶ Flow data provided as relative trips as opposed to absolute counts
- ▶ Convert relative flows to trips using HH travel survey?

CHTS



Google AAT



Relative Flow Conversion Model



$$PT_{odt} = \beta_t(RF_{odt}) + \epsilon; \quad 0 \leq t \leq 23$$

PT_{odt} : Person trips between o-d for hour-of-day t

β_t : Coefficient of AAT relative flow for t

RF_{odt} : Avg AAT flow between o-d for hour-of-day t

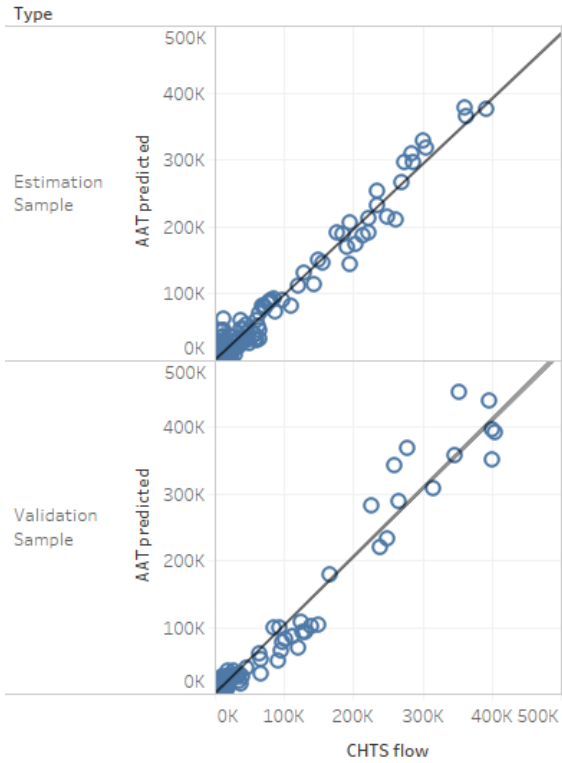
ϵ : Error term

- ▶ No geographic constants applied
- ▶ District- and County-level regression models estimated
- ▶ 20% sample used for validation

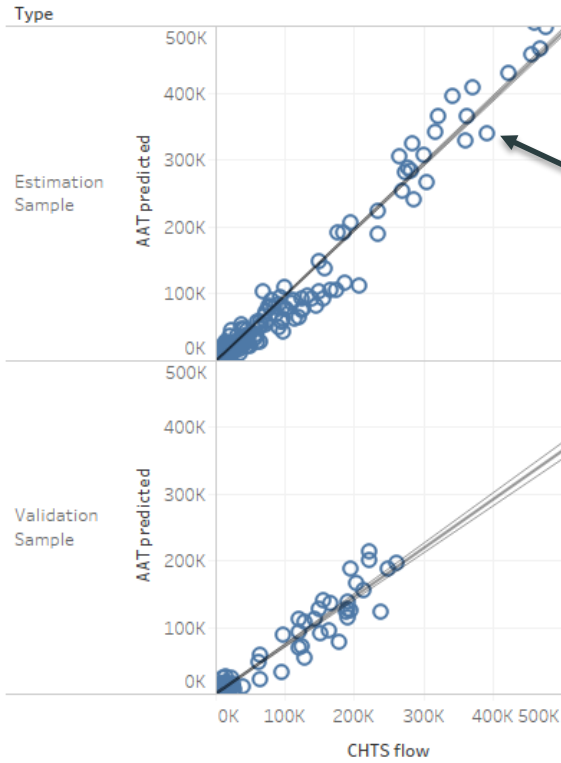
Predicted vs Observed



District Model

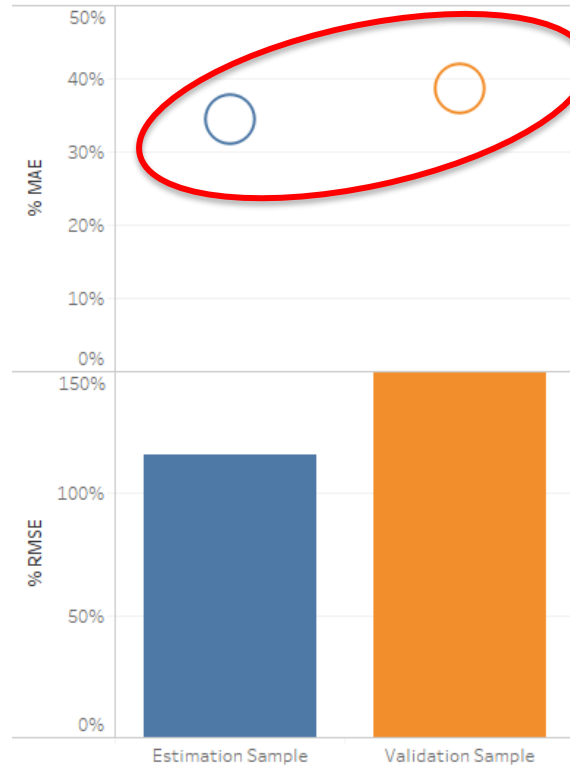


County Model

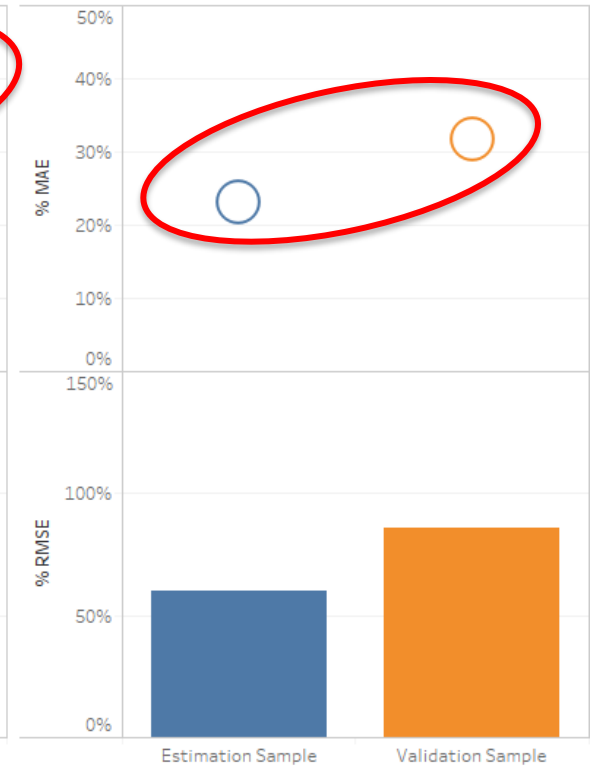


Origin-Destination-Hour

District Model



County Model



Summary



- ▶ **AAT relative flow magnitudes correlated with actual trips**
- ▶ **AAT geographic coverage significantly higher**
- ▶ **Simple linear regression model may be used for conversion**
- ▶ **Could support measuring longitudinal variation**
- ▶ **Further studies**
 - ▶ **Better/smooth survey data**
 - ▶ **Compare with cell CDR data**