

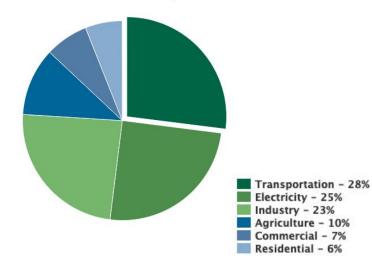
Illustration via iStock

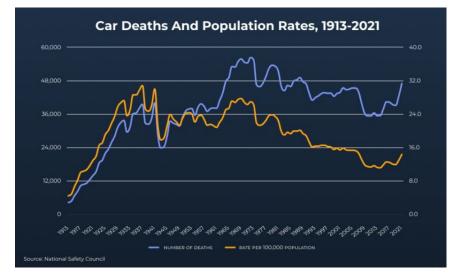
#### Investments in Public Transit are Crucial, but Costly

Greenhouse gas emissions

**Traffic deaths & injuries** 

#### 2022 U.S. GHG Emissions by Sector





# **Our Goals:**

- Forecast transit ridership for U.S. cities
- Identify key factors for improving ridership



#### The National Transit Database



- Records the financial, operating, and asset condition of transit systems throughout the United States
- Our dataset: 182 cities across the US with statistics from 1991 to 2023
- Trained models on pre-2019 data, and tested on 2019 data
- Discarded post-2019 data due to COVID-19 ridership slump

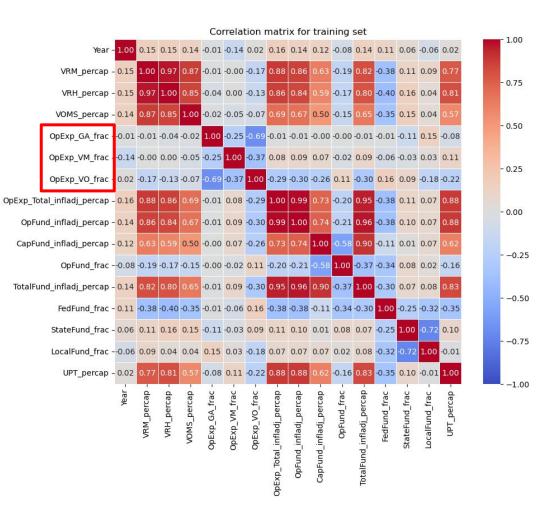
- Statistics on vehicles, operational expenses, funding sources, and ridership
- Removed features with strong multicollinearity

Correlation matrix for training set														- 1.00					
	Year -	1.00	0.15	0.15	0.14	-0.01	-0.14	0.02	0.16	0.14	0.12	-0.08	0.14	0.11	0.06	-0.06	0.02		- 1.00
	VRM_percap -	0.15	1.00	0.97	0.87	-0.01	-0.00	-0.17	0.88	0.86	0.63	-0.19	0.82	-0.38	0.11	0.09	0.77		0.75
	VRH_percap	0.15	0.97	1.00	0.85	-0.04	0.00	-0.13	0.86	0.84	0.59	-0.17	0.80	-0.40	0.16	0.04	0.81		- 0.75
VOMS_percap		0.14	0.87	0.85	1.00	-0.02	-0.05	-0.07	0.69	0.67	0.50	-0.15		-0.35	0.15	0.04	0.57		
	OpExp_GA_frac	-0.01	-0.01	-0.04	-0.02	1.00	-0.25	-0.69	-0.01	-0.01	-0.00	-0.00	-0.01	-0.01	-0.11	0.15	-0.08		- 0.50
	OpExp_VM_frac ·	-0.14	-0.00	0.00	-0.05	-0.25	1.00	-0.37	0.08	0.09	0.07	-0.02	0.09	-0.06	-0.03	0.03	0.11		
	OpExp_VO_frac ·	0.02	-0.17	-0.13	-0.07	-0.69	-0.37	1.00	-0.29	-0.30	-0.26	0.11	-0.30	0.16	0.09	-0.18	-0.22		- 0.25
OpExp_Total_infladj_percap		0.16	0.88	0.86	0.69	-0.01	0.08	-0.29	1.00	0.99	0.73	-0.20	0.95	-0.38	0.11	0.07	0.88		
OpFund_infladj_percap		0.14	0.86	0.84	0.67	-0.01	0.09	-0.30	0.99	1.00	0.74	-0.21	0.96	-0.38	0.10	0.07	0.88		- 0.00
CapFund_infladj_percap		0.12	0.63	0.59	0.50	-0.00	0.07	-0.26	0.73	0.74	1.00	-0.58	0.90	-0.11	0.01	0.07	0.62		
OpFund_frac		-0.08	-0.19	-0.17	-0.15	-0.00	-0.02	0.11	-0.20	-0.21	-0.58	1.00	-0.37	-0.34	0.08	0.02	-0.16		0.25
TotalFund_infladj_percap		0.14	0.82	0.80	0.65	-0.01	0.09	-0.30	0.95	0.96	0.90	-0.37	1.00	-0.30	0.07	0.08	0.83		
	FedFund_frac	0.11	-0.38	-0.40	-0.35	-0.01	-0.06	0.16	-0.38	-0.38	-0.11	-0.34	-0.30	1.00	-0.25	-0.32	-0.35		0.50
	StateFund_frac	0.06	0.11	0.16	0.15	-0.11	-0.03	0.09	0.11	0.10	0.01	0.08	0.07	-0.25	1.00	-0.72	0.10		
	LocalFund_frac	-0.06	0.09	0.04	0.04	0.15	0.03	-0.18	0.07	0.07	0.07	0.02	0.08	-0.32	-0.72	1.00	-0.01		0.75
	UPT_percap -	0.02	0.77	0.81	0.57	-0.08	0.11	-0.22	0.88	0.88	0.62	-0.16	0.83	-0.35	0.10	-0.01	1.00		
		Year -	VRM_percap -	VRH_percap -	VOMS_percap -	OpExp_GA_frac -	OpExp_VM_frac -	OpExp_VO_frac -	OpExp_Total_infladj_percap -	OpFund_infladj_percap -	CapFund_infladj_percap -	OpFund_frac -	TotalFund_infladj_percap -	FedFund_frac -	StateFund_frac -	LocalFund_frac -	UPT_percap -		1.00

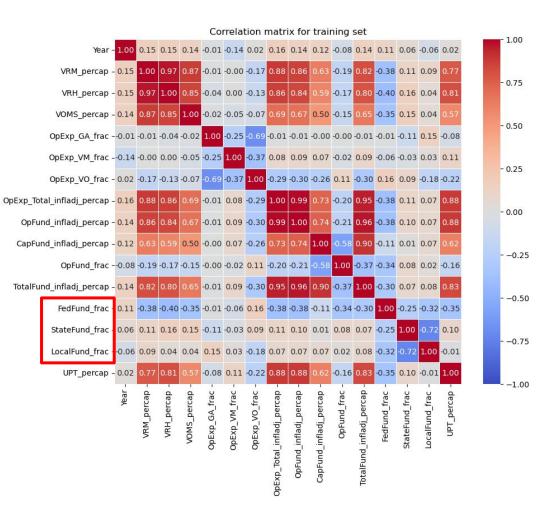
How many hours are all vehicles collecting revenue (VRH)?

Correlation matrix for training set														1.00		
Year - <mark>1.00</mark>	0.15	0.15 0.14	-0.01	-0.14	0.02	0.16	0.14	0.12	-0.08	0.14	0.11	0.06	-0.06	0.02		1.00
VRM_percap - 0.15	1.00	0.97 0.83	-0.01	-0.00	-0.17	0.88	0.86		-0.19	0.82	-0.38	0.11	0.09	0.77		
VRH_percap - 0.15	0.97	1.00 0.85	-0.04	0.00	-0.13	0.86	0.84		-0.17	0.80	-0.40	0.16	0.04	0.81	-	0.75
VOMS_percap - 0.14	0.87	0.85 1.00	-0.02	-0.05	-0.07	0.69	0.67	0.50	-0.15		-0.35	0.15	0.04	0.57		
OpExp_GA_frac0.01	-0.01 -	-0.04 -0.0	2 1.00	-0.25	-0.69	-0.01	-0.01	-0.00	-0.00	-0.01	-0.01	-0.11	0.15	-0.08	-	0.50
OpExp_VM_frac0.14	-0.00	0.00 -0.0	5 -0.25	1.00	-0.37	0.08	0.09	0.07	-0.02	0.09	-0.06	-0.03	0.03	0.11		
OpExp_VO_frac - 0.02	-0.17 -	-0.13 -0.0	7 -0.69	-0.37	1.00	-0.29	-0.30	-0.26	0.11	-0.30	0.16	0.09	-0.18	-0.22	-	0.25
OpExp_Total_infladj_percap - 0.16	0.88	0.86 0.69	-0.01	0.08	-0.29	1.00	0.99	0.73	-0.20	0.95	-0.38	0.11	0.07	0.88		
OpFund_infladj_percap - 0.14	0.86	0.84 0.6	-0.01	0.09	-0.30	0.99	1.00	0.74	-0.21	0.96	-0.38	0.10	0.07	0.88	-	0.00
CapFund_infladj_percap - 0.12	0.63	0.59 0.50	0.00	0.07	-0.26	0.73	0.74	1.00	-0.58	0.90	-0.11	0.01	0.07	0.62		
OpFund_frac0.08	-0.19 -	-0.17 -0.1	5 -0.00	-0.02	0.11	-0.20	-0.21	-0.58	1.00	-0.37	-0.34	0.08	0.02	-0.16	-	-0.25
TotalFund_infladj_percap - 0.14	0.82	0.80 0.65	-0.01	0.09	-0.30	0.95	0.96	0.90	-0.37	1.00	-0.30	0.07	0.08	0.83		
FedFund frac - 0.11	-0.38 -	-0.40 -0.3	5 -0.01	-0.06	0.16	-0.38	-0.38	-0.11	-0.34	-0.30	1.00	-0.25	-0.32	-0.35	-	-0.50
StateFund frac - 0.06	0.11	0.16 0.15	-0.11	-0.03	0.09	0.11	0.10	0.01	0.08	0.07	-0.25	1.00	-0.72	0.10		
LocalFund frac0.06	0.09	0.04 0.04	0.15	0.03	-0.18	0.07	0.07	0.07	0.02	0.08	-0.32	-0.72	1.00	-0.01	-	-0.75
UPT percap - 0.02				0.11							-0.35					
	10 14		1	- 1	- 1	-	e ye w	-		100	,	1			-	-1.00
fear	VRM_percap	VRH_percap	OpExp_GA_frac	OpExp_VM_frac	DpExp_VO_frac	dj_perca	dj_perca	dj_perca	OpFund_frac	dj_perca	FedFund_frac	StateFund_frac	LocalFund_frac	UPT_percap		
	VRI	VRI	OpExp	OpExp	OpExp	OpExp_Total_infladj_percap	OpFund_infladj_percap	CapFund_infladj_percap	Jdo	TotalFund_infladj_percap	FedF	StateF	LocalF	-U		

What fraction of funds are spent on administration, vehicle maintenance, & vehicle operation?

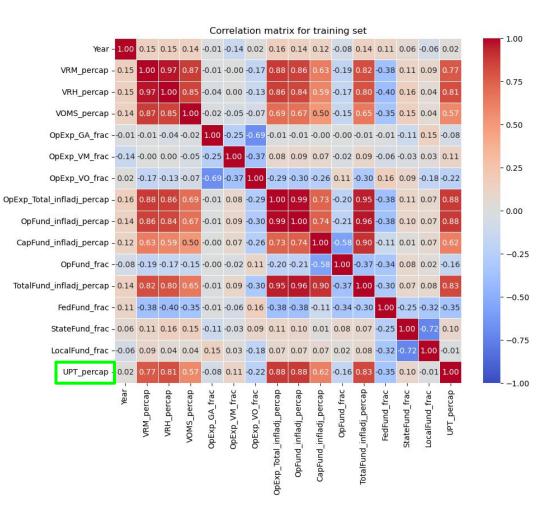


What fraction of funding comes from federal, state, and local governments?



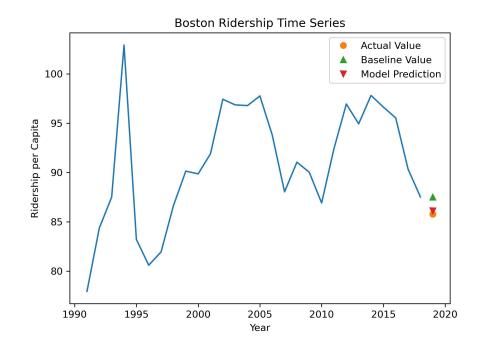
Forecast variable:

Ridership is measured by annual Unlinked Passenger Trips (UPT) per capita



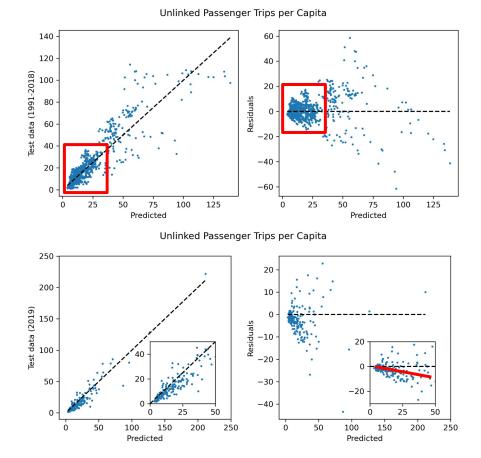
#### A City by City Model

- Unique model trained for each city
- Lasso regression eliminates extraneous features
- Example: Boston, MA
  - Features with non-zero coefficients were Vehicle Maintenance Expenses, Federal, and State Funding.
  - Our model provides a better prediction than the baseline
- Bespoke model out-performs a naive forecast for only 36/182 cities



#### **All-Cities Model**

- Trained on 1991-2018 data for 80% of cities
- Cross-validation: XGBoost out-performed baseline and multiple linear regression
- Tests well on pre-2019 data
  Except high-ridership cities
- Residuals of 2019 data reveal un-modeled trend



#### **Insights and Shortcomings**

- The City by City Model provided useful insights into where transit agency can dedicate resources.
- However, it performed worse than a naive forecast for most cities.
- We may be missing important factors for understanding ridership.

- The All-City Model performed better due to a much larger training set.
- Additionally, it showed trends across different cities.
- However, large residuals exist for some larger cities.

# Future Work

 Develop unique models for cities of different size categories

**King County** 

-90,693

CAPITALIST

(Seattle)

+335,884

- Better account for time lag between features and ridership (e.g., capital expenses)
- Investigate population changes, local developments, and other features not included in the NTD

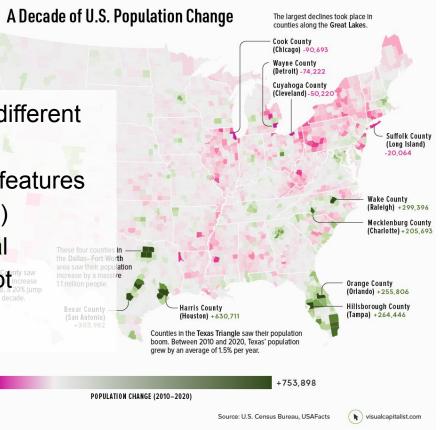




Illustration via iStock