Keeping Up: Non-traditional Traffic Monitoring Development

October 31st 2018

Laura Schewel
CEO, StreetLight Data
laura.schewel@streetlightdata.com

STREETLIGHTDATA

Agenda for Today's Seminar

I. Introduction to StreetLight Volume: 2017 AADT Estimates

II. Quick Demo

III. Data Sources

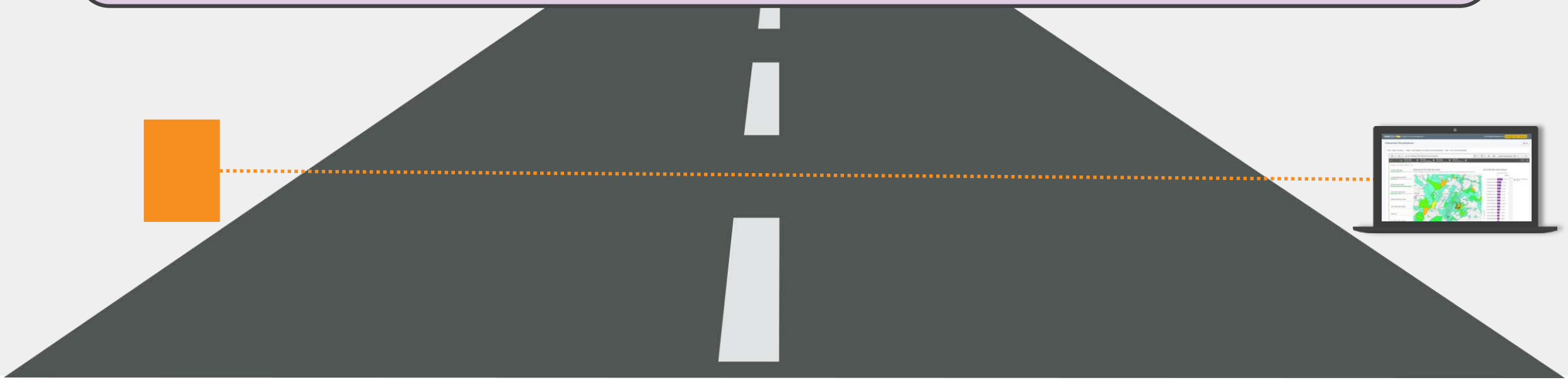
IV. Questions?

Introducing StreetLight Volume: 2017 AADT Estimates

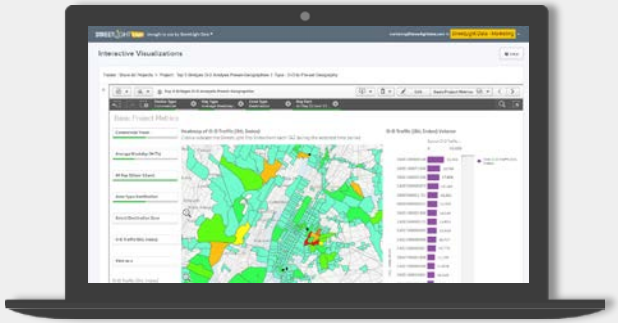
STREETLIGHTDATA

StreetLight Volume: 2017 AADT – What Is It?

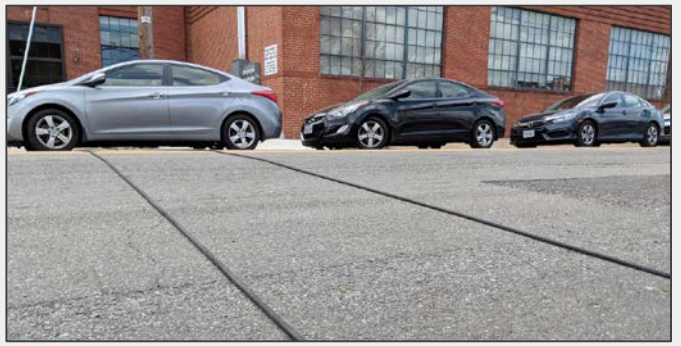
Average Annual Daily Traffic *Counts*



We Believe Our New AADT Estimates Are the Best Alternative to Temporary and Modeled AADT



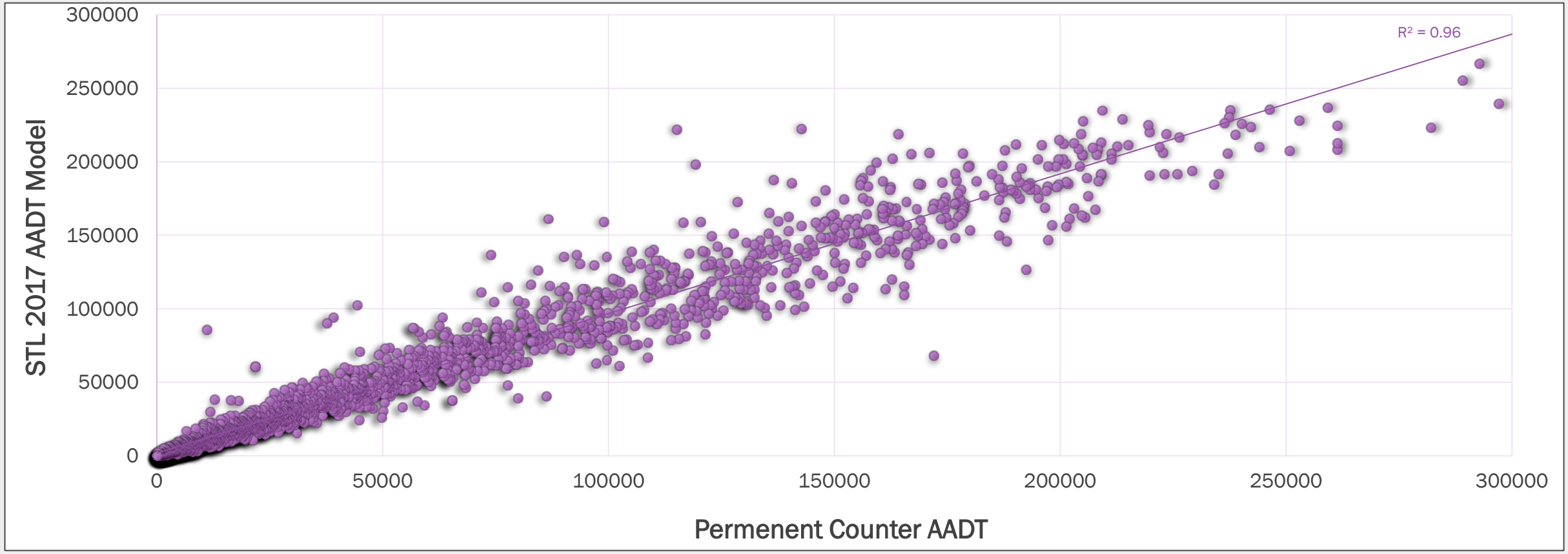
- No staff in harm's way
- Cost-effective
- Available in minutes
- 365 days of real-world data
- **As accurate to more accurate than temporary/modeled counts**



- Staff in field in harms way
- Expensive
- Time-intensive data collection and processing
- 2 to 7 days of real-world data

These Are the Results of Our Validation Work

StreetLight 2017 AADT for Test Data compared to Permanent Counter AADT. R2 is 0.96. No outliers were removed.



The Results from Our National Validation Tests

Key Results from National Validation Test –Absolute Error

AADT Range	# of Segments	Target Abs. Error	StreetLight Algorithm Mean Abs. Error	Delta to Target (positive means "better than target")
50,000+	795	12%	12.7%	-1%
25,000- 49,999	386	16%	15.7%	0%
10,000 – 24,999	509	20%	20.6%	-1%
5,000- 9,999	350	20%	23.5%	-4%
0 – 4,999	564	Not available - too few in comparison paper. See Table 1b.	43.3%	NA

Key Results from National Validation Test –RMSE

AADT Range	# of segments	Target RMSE as % of Average AADT	StreetLight Algorithm's RMSE as % Average AADT	Delta to Target (positive means "better than target")
50,000+	795	20%	15.8%	4%
25,000- 49,999	386	25%	20.8%	4%
10,000 – 24,999	509	28%	31.4%	-3%
5,000- 9,999	350	39%	31.5%	7%
2500 – 4,999	270	44%	36.1%	8%
0 – 2,499	294	68%	58.8%	9%

Quick Demo – How to Derive AADT in *StreetLight InSight*

STREETLIGHTDATA

Data Sources

STREETLIGHTDATA

We Use Six Unique Data Sources for Our StreetLight Volume: 2017 AADT Estimates

1

Big Data Input #1

Full Year of 2017 Location-Based Services Data

2

Big Data Input #2

Full Year of 2017 Navigation-GPS Data

3

Contextual Input #1

US Census Data for Normalization

United States Census Bureau

4

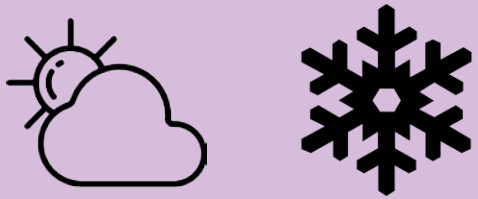
Contextual Input #2

Open Street Maps Data

5

Contextual Input #3

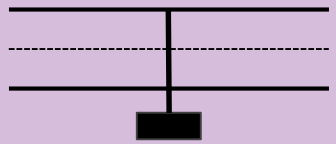
Weather Data



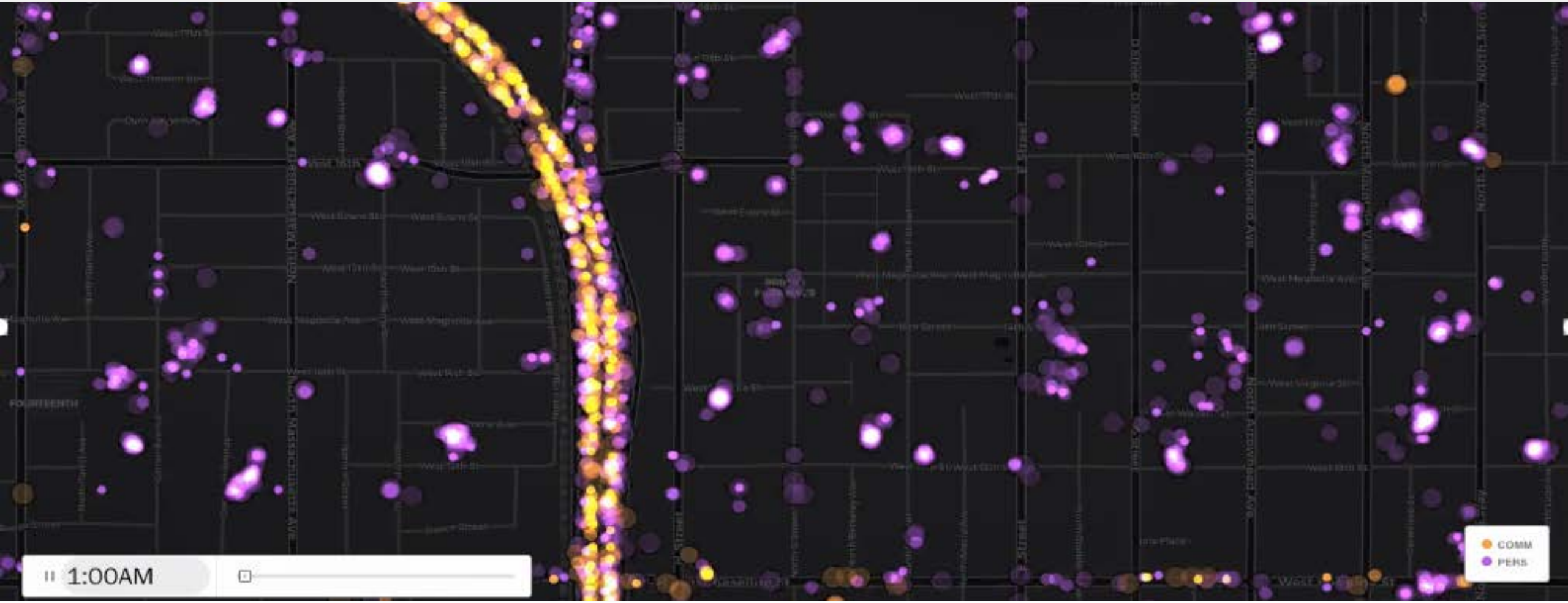
6

Permanent Counters

Data from 2,000+ Permanent Counters



Our Big Data Resources: Location-Based Services and Navigation-GPS Data



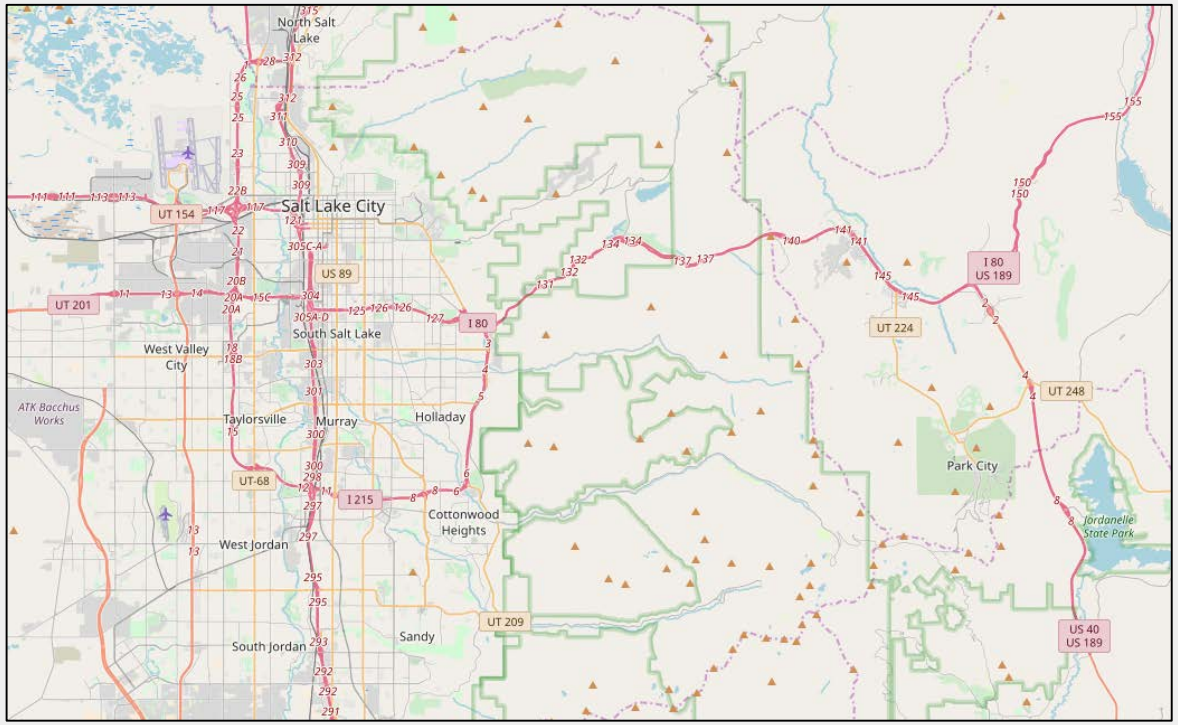
Our Big Data Resources: Location-Based Services and Navigation-GPS Data

Navigation-GPS Data: Created by Connected Trucks & Cars	
Spatial Precision	~5 meters
Frequency of Data Pings	Regularly; every 1 sec – 1 min
Type of Trip	Differentiates personal and commercial trips – ideal for truck studies
Sample Size	Penetration rate varies by region – but much smaller than LBS. ~1% - 4% for personal, 12% trucks.

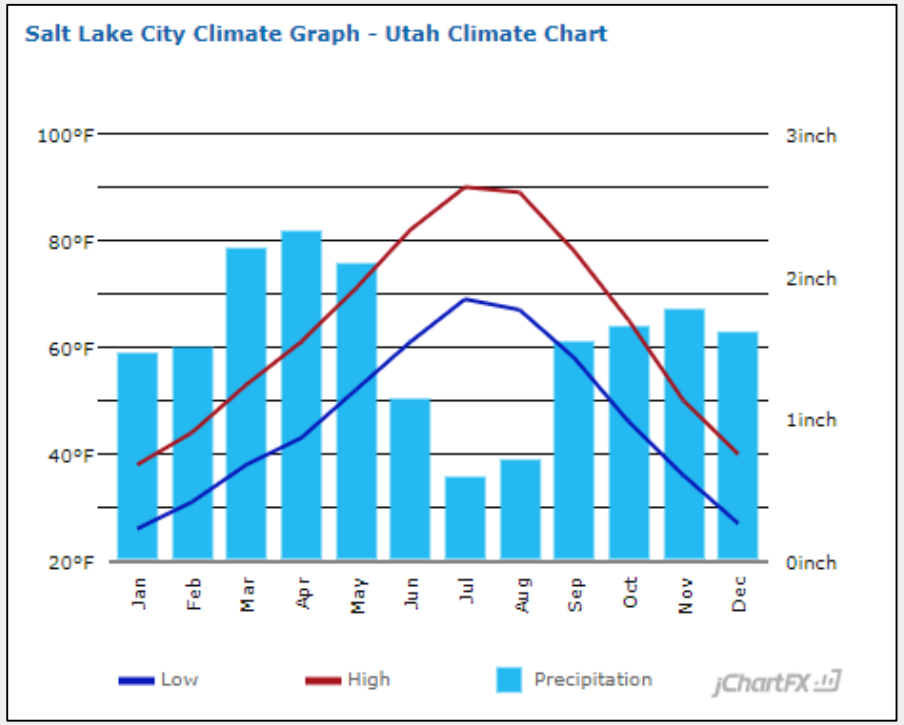
Location-Based Services (LBS) Data: Created by Smart Phone Apps	
Spatial Precision	~5 meters – 25 meters
Frequency of Data Pings	Variable; usually triggered by location change
Type of Trip	Personal
Sample Size	~23% of US and CA adult population (~65M devices in our database)

We Used Two Different Contextual Data Sources to Account for Roadway and Environmental Factors

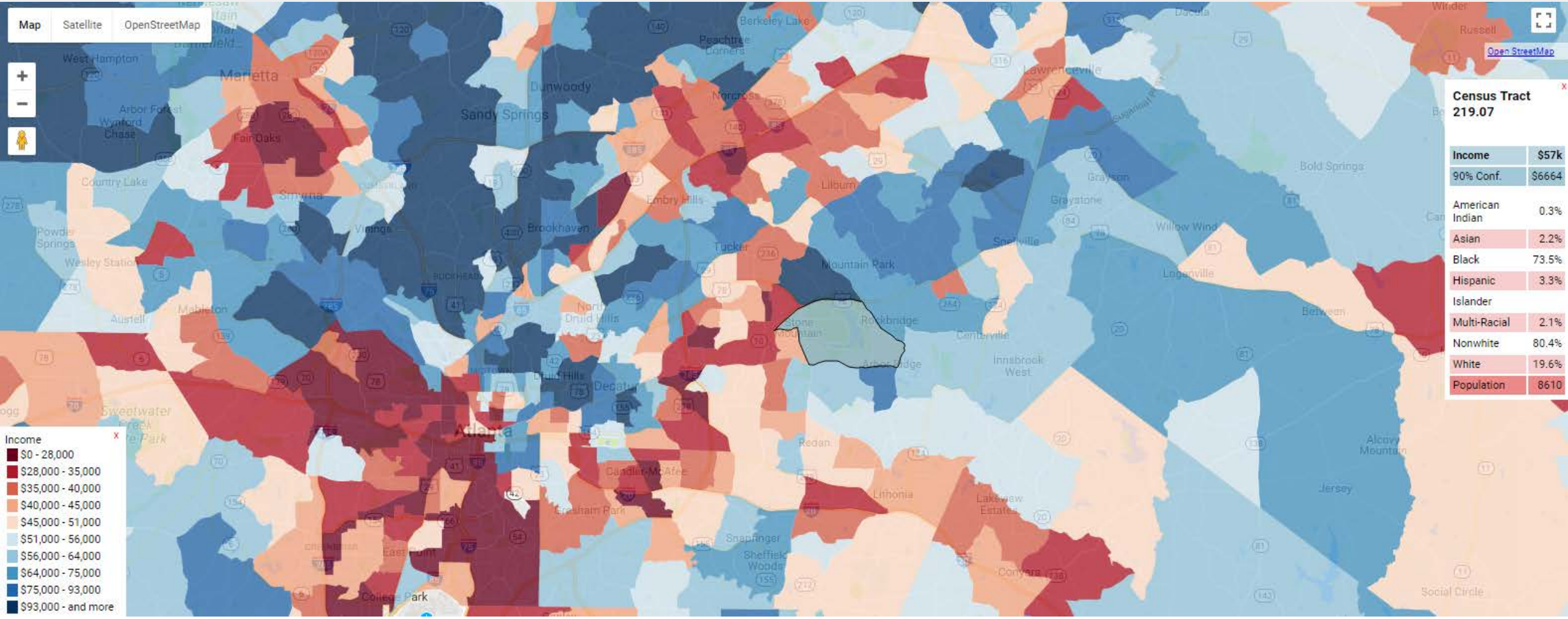
A Look into Open Street Maps: Salt Lake City, UT & Surroundings



A Look into Weather Data: Precipitation & Temperature in Salt Lake City, UT



US Census Data – Our Third Contextual Resource – Was Used for Normalization of LBS Trips

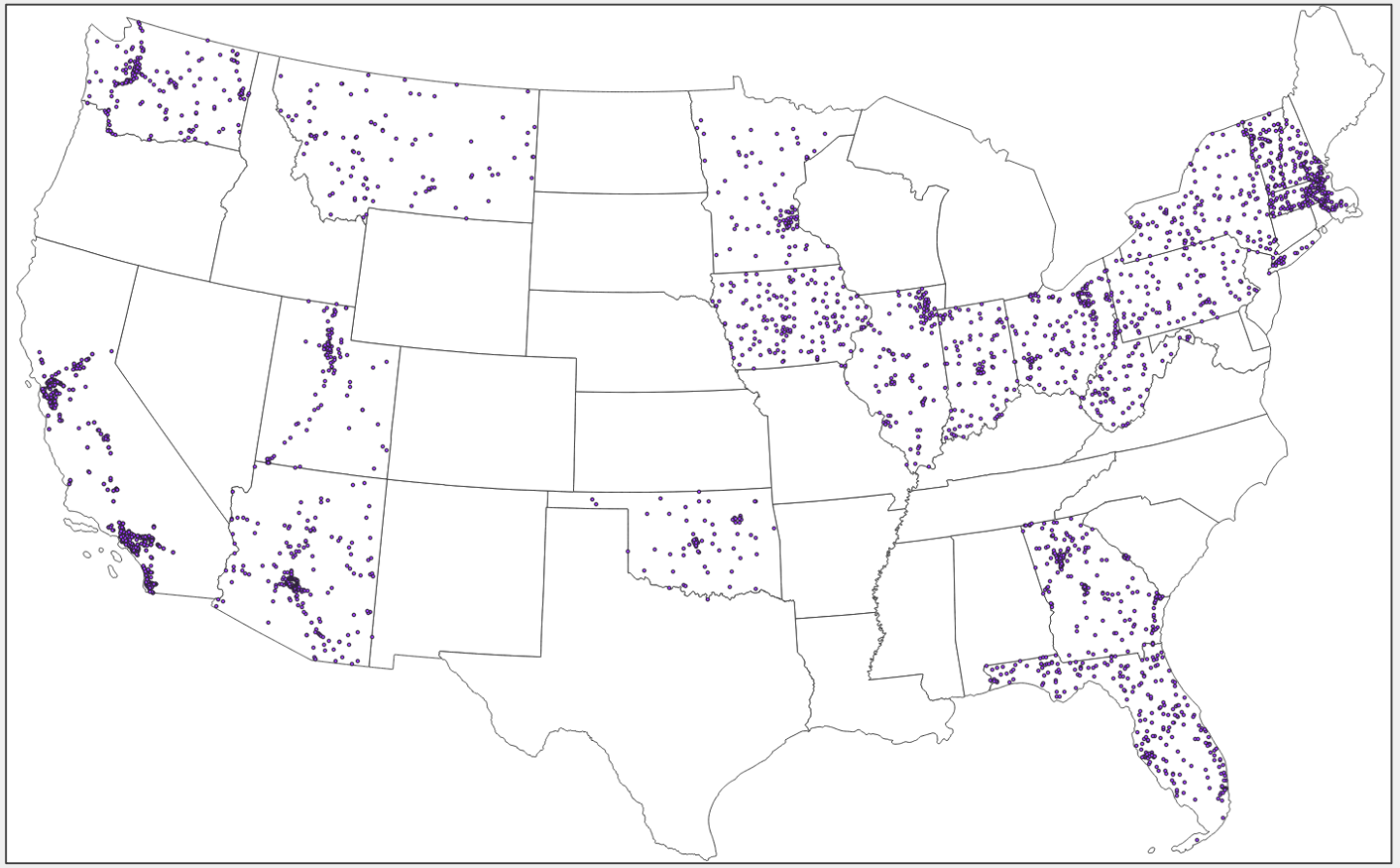


Our Data Resources from 2,000+ Permanent Count Locations Were Critical to Algorithm Development

Locations and AADT Distribution of the 2,605 Permanent Counters

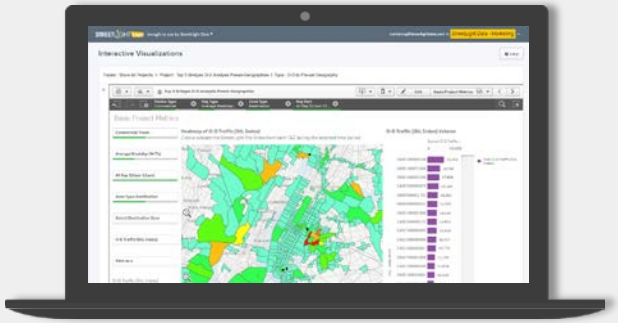
State	# of Counters	State	# of Counters
AZ	232	NY	144
FL	243	NH	65
GA	181	OH	146
ID	116	OK	68
IN	90	CA	272
IA	147	PA	90
MA	193	UT	108
MN	84	VT	82
MT	97	WA	175

AADT Range	# of Counters
50,000+	795
25,000- 49,999	386
10,000 – 24,999	509
5,000 - 9,999	350
2500 – 4,999	270
0 – 2,499	294

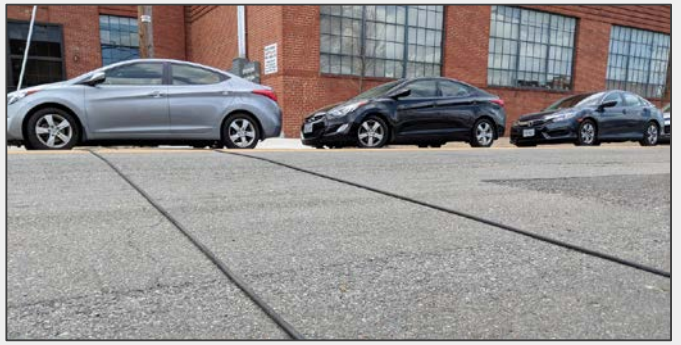


STREETLIGHTDATA

Full Circle: Our 2017 AADT Estimates Are Better than Temporary and Modeled AADT Counts



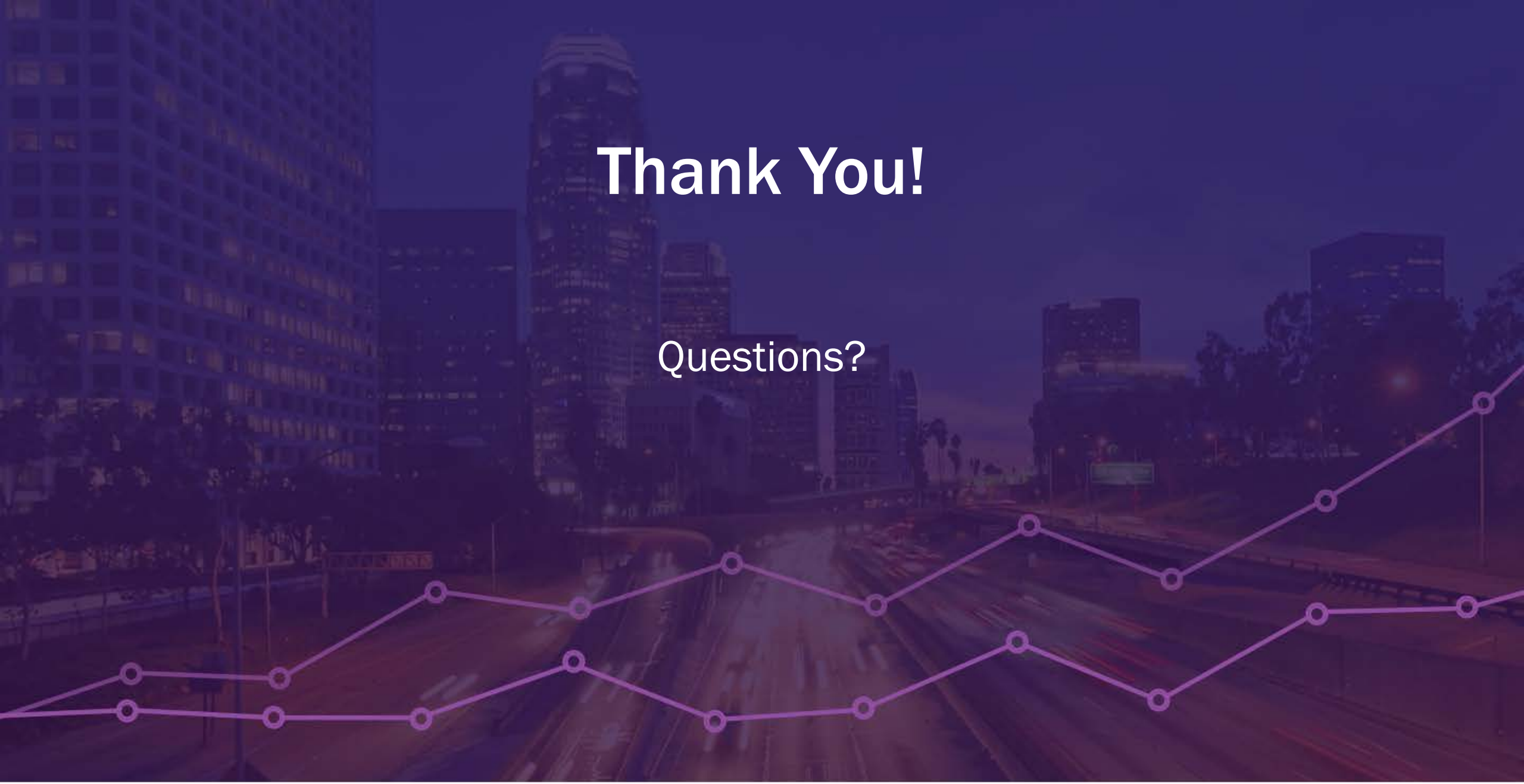
- No staff in harm's way
- Cost-effective
- Available in minutes
- 365 days of real-world data
- **As accurate to more accurate than temporary/modeled counts**



- Staff in field in harms way
- Expensive
- Time-intensive data collection and processing
- 2 to 7 days of real-world data

Big Data Can Do More Than Just Replace:
The Wheel of Putting Big Data to Work

STREETLIGHTDATA



Thank You!

Questions?

STREETLIGHTDATA

Goal #1: Develop Estimates that Are Better Than Temporary Counts

Equation for Root Mean Square Error (RMSE) as a Percent of AADT

$$MAPE = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} 100 * \left| \frac{AADT_{Estimate(i)} - AADT_{Permanent\ counter(i)}}{AADT_{Permanent\ counter(i)}} \right|$$

Source: Gadda, S., A. Mangoon, and K. Kockelman. "Estimates of AADT: Quantifying the Uncertainty." 11th World Conference on Transport Research. Location: Berkeley CA, United States. Date: 2007-6-24 to 2007-6-28

Acceptable Range Per University of Texas Researchers:

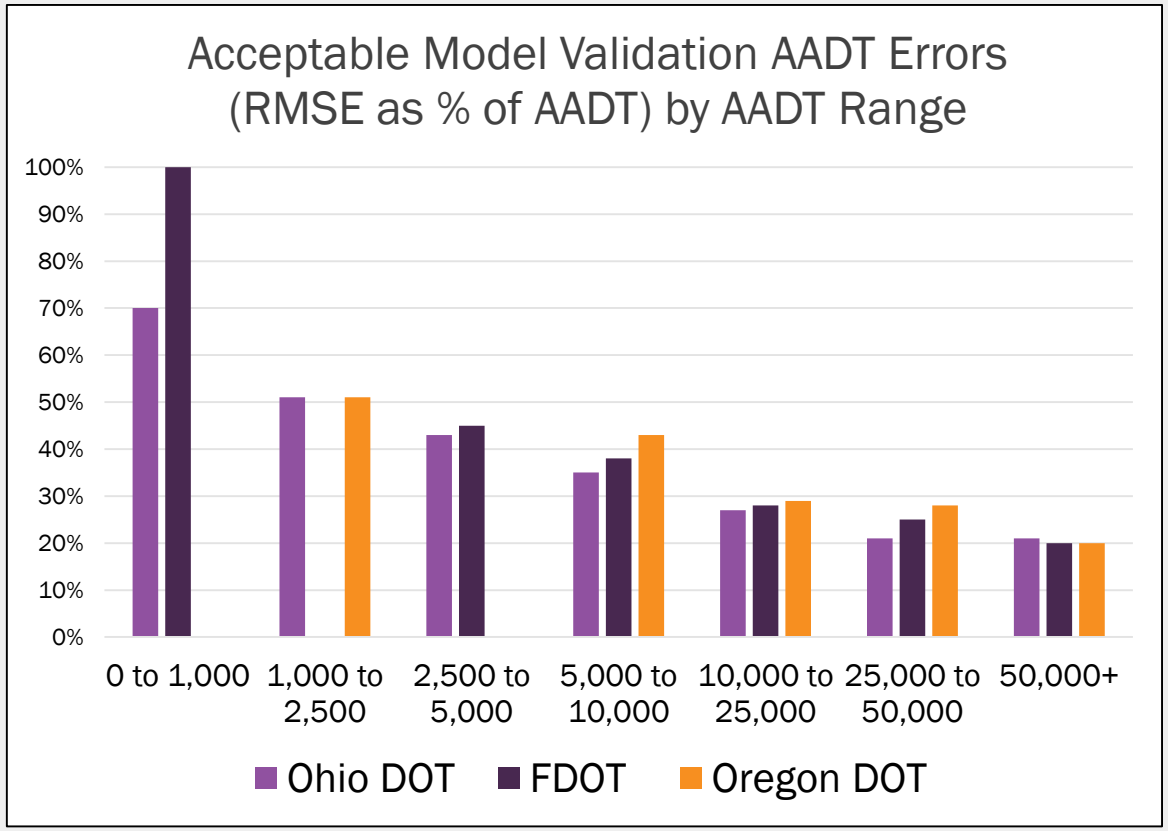
4.9% – 83%

Our Overall Targets:

20% for lower AADT bins

12% for higher AADT bins

Goal #2: Develop Estimates that Are Better Than Counts Derived from Expansion Models



Equation for Root Mean Square Error (RMSE) as a Percent of AADT

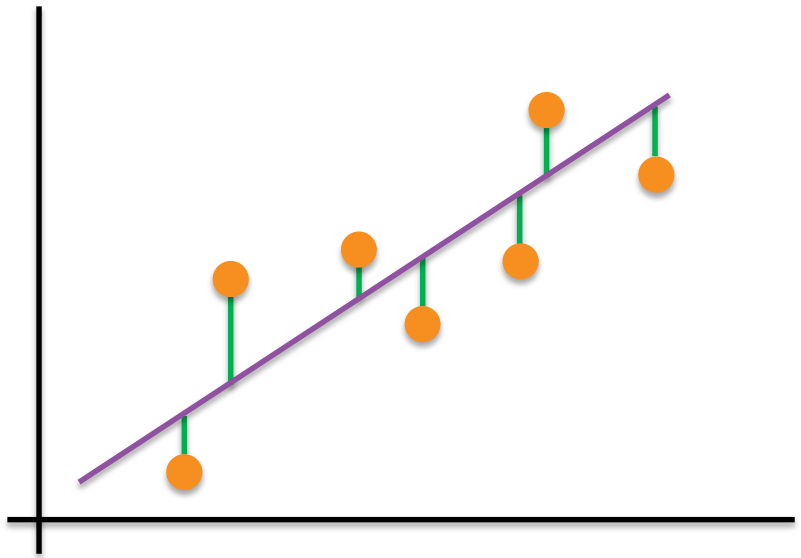
$$RMSE\ as\ \% = 100 * \frac{\sqrt{\left(\frac{1}{n}\right) * (Traditional\ AADT - Big\ Data\ AADT)^2}}{\left(\frac{1}{n} * \sum_{i=1}^{n}(Traditional\ AADT)\right.}$$

Source – Figure 9.8 in Travel Model Improvement Program, "Travel Model Validation and Reasonableness Checking Manual Second Edition." September 24, 2010.

Algorithm Development

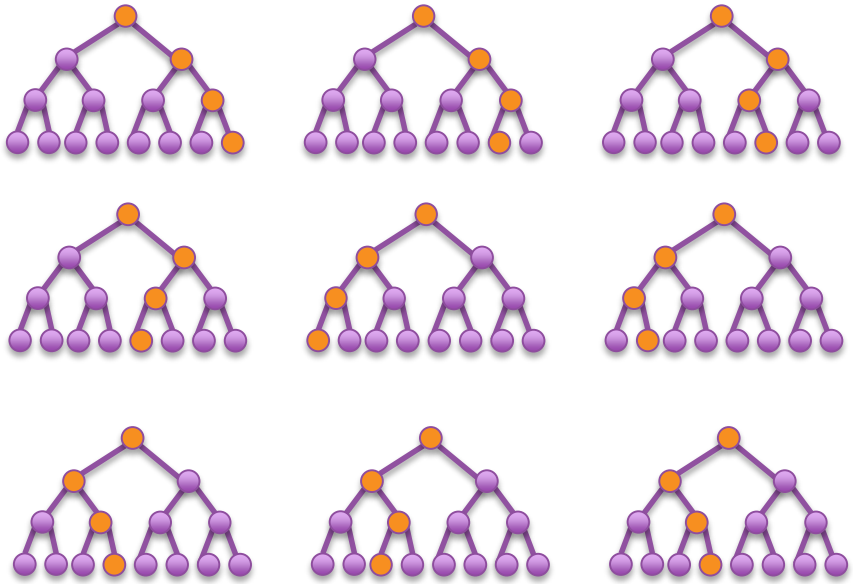
Our Second Step: Selecting the Algorithm

Ordinary Least Squares (OLS)



VS.

Random Forest



We Selected the Random Forest Algorithm for Our 2017 AADT Estimates Model

- **More accurate:**
  Better at handling unusual roads
- **Final Result:**
  We built a 12-feature Random Forest model