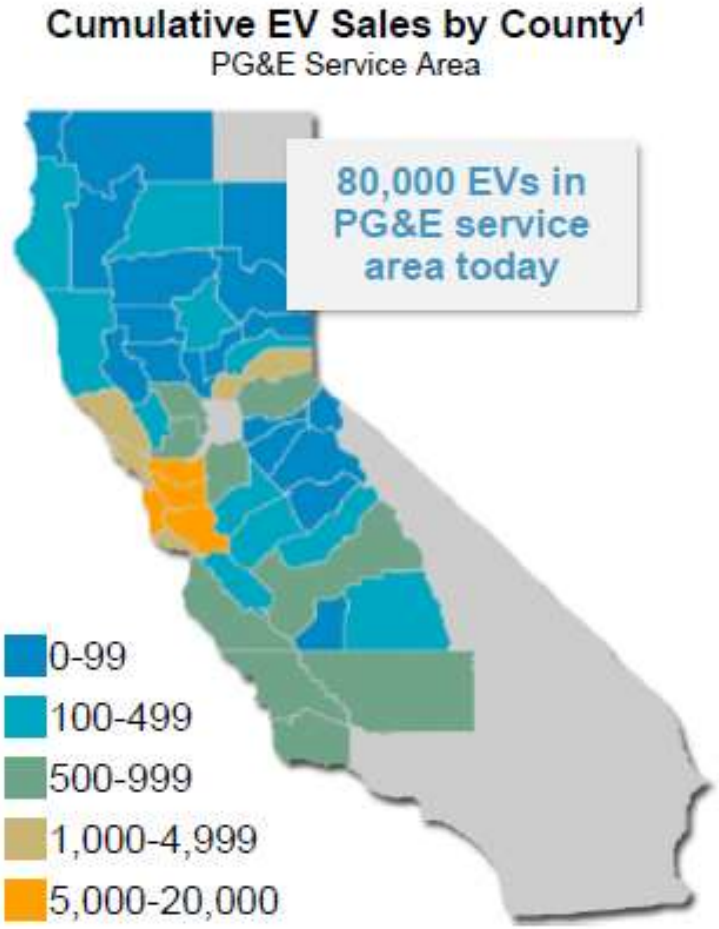
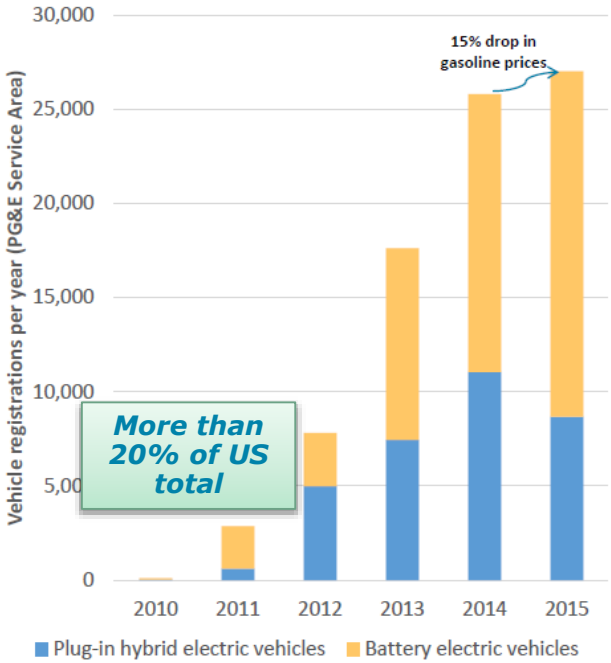
EV Owner Identification Study

April 2016

Erin Boyd on behalf of Andrew Lee
Measurement and Evaluation
Pacific Gas & Electric Co.

PG&E has more electric vehicles plugging in than any utility in the US





Governor Jerry Brown set a California goal of 1.5 million zero-emission vehicles by 2025, with the infrastructure to support one million EVs by 2020

Electric vehicle rate options

			Peak	Partial Peak	Off Peak
EV Rates	EV TOU A (Whole House)	Summer	0.44402	0.24156	0.11466
		Winter	0.31228	0.19043	0.11742
	EV TOU B (Separately Metered)	Summer	0.43755	0.23832	0.11419
		Winter	0.30539	0.18699	0.11692

TOU rate options comparison

			Peak	Partial Peak	Off Peak
EV Rates	EV TOU A (Whole House)	Summer	0.44402	0.24156	0.11466
		Winter	0.31228	0.19043	0.11742
	EV TOU B (Separately Metered)	Summer	0.43755	0.23832	0.11419
		Winter	0.30539	0.18699	0.11692
Non-EV Rates	TOU A	Summer	0.40227		0.32669
		Winter	0.28430		0.27000
	TOU B	Summer	0.35600		0.25294
		Winter	0.21854		0.19974

Motivation for study

- Would like to measure impact of TOU rates on charging behavior of EV customers
- EV owners not on TOU rate best candidates for developing a comparison group for measurement – although self-selection bias, matching methods combined with models to account for selection bias can be used to produce valid impact estimates given large enough sample sizes

Objective and approach

Objective: Identify a control group of EV owners to use for a load impact analysis of the EV TOU rate

Objective and approach

Objective: Identify a control group of EV owners to use for a load impact analysis of the EV TOU rate

Due to universal deployment of AMI in PG&E territory, possible to search for large EV charging signals in load data to identify customers who are likely to own an EV

Objective and approach

Objective: Identify a control group of EV owners to use for a load impact analysis of the EV TOU rate

Due to universal deployment of AMI in PG&E territory, possible to search for large EV charging signals in load data to identify customers who are likely to own an EV

Create a pool of customers that consists of known EV owners and non-EV owners and divide them into a training set and test set

→

Develop a set of predictive algorithms based on the training set

→

Run the algorithms on the test set and identify the best performing one

→

Run algorithm on population of non-EV rate customers and distribute surveys to those predicted to have an EV to confirm ownership

Objective and approach

Objective: Identify a control group of EV owners to use for a load impact analysis of the EV TOU rate

Due to universal deployment of AMI in PG&E territory, possible to search for large EV charging signals in load data to identify customers who are likely to own an EV

Create a pool of customers that consists of known EV owners and non-EV owners and divide them into a training set and test set → Develop a set of predictive algorithms based on the training set → Run the algorithms on the test set and identify the best performing one → Run algorithm on population of non-EV rate customers and distribute surveys to those predicted to have an EV to confirm ownership

Repeat process using higher quality data

A representative sample of the PG&E population with known EV and non-EV customers does not exist

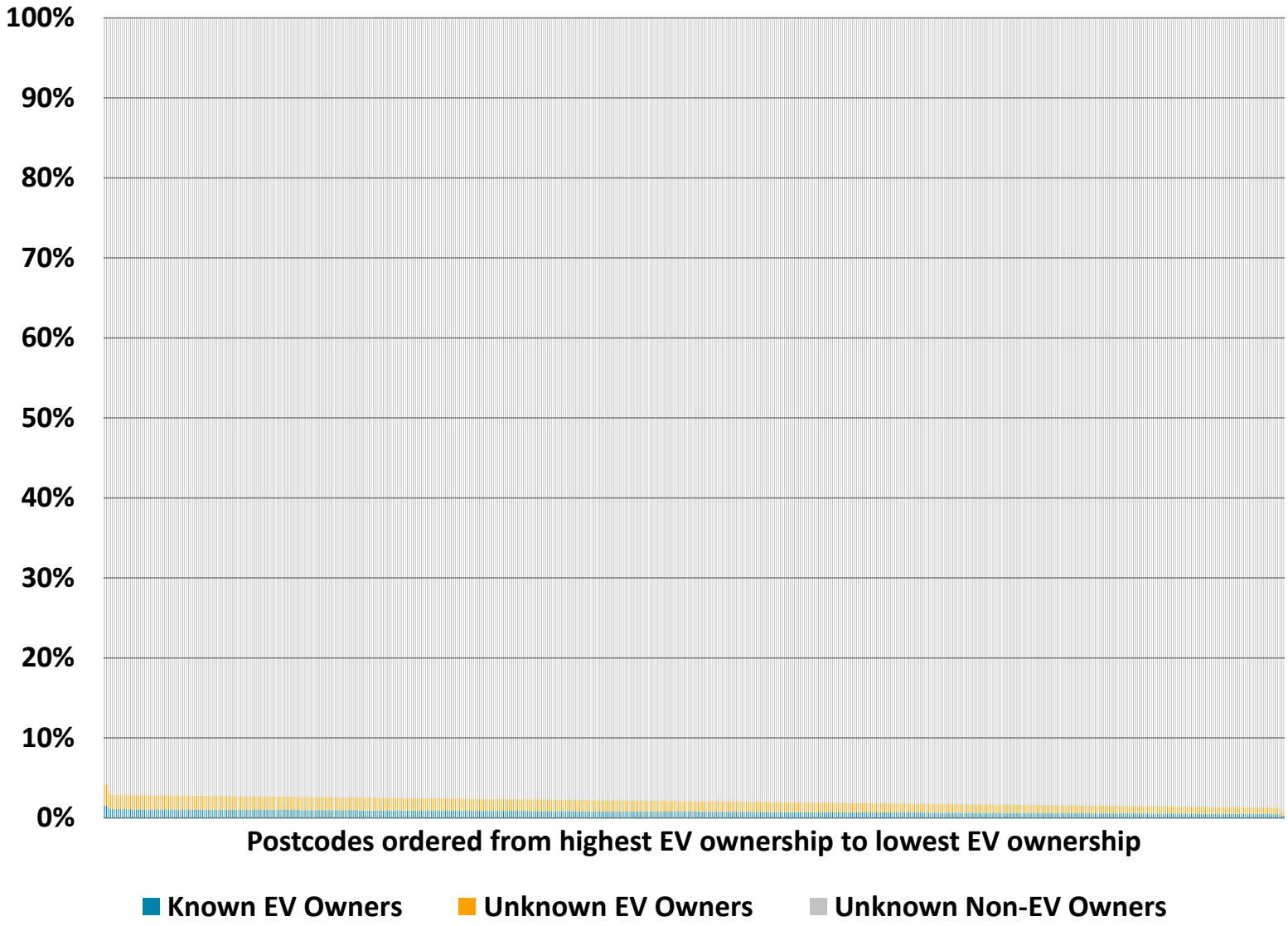
- Customers who purchase EV under no obligation to disclose purchase to utility or sign up for a TOU rate
- Only subset of EV population is known – estimate approximately 1/3 of EV population on EV TOU rate. . .*likely not representative of EV customers not on a TOU rate due to unobservable characteristics related to decision to enroll on EV TOU and effects of TOU rate on load*
- Even if EV TOU customers representative of all EV customers in general population, a sample of customers that definitely did not own EVs necessary to estimate any classification algorithm
- Remaining 2/3 EV owners unidentified in general residential population of approximately 6.4 million residential customers

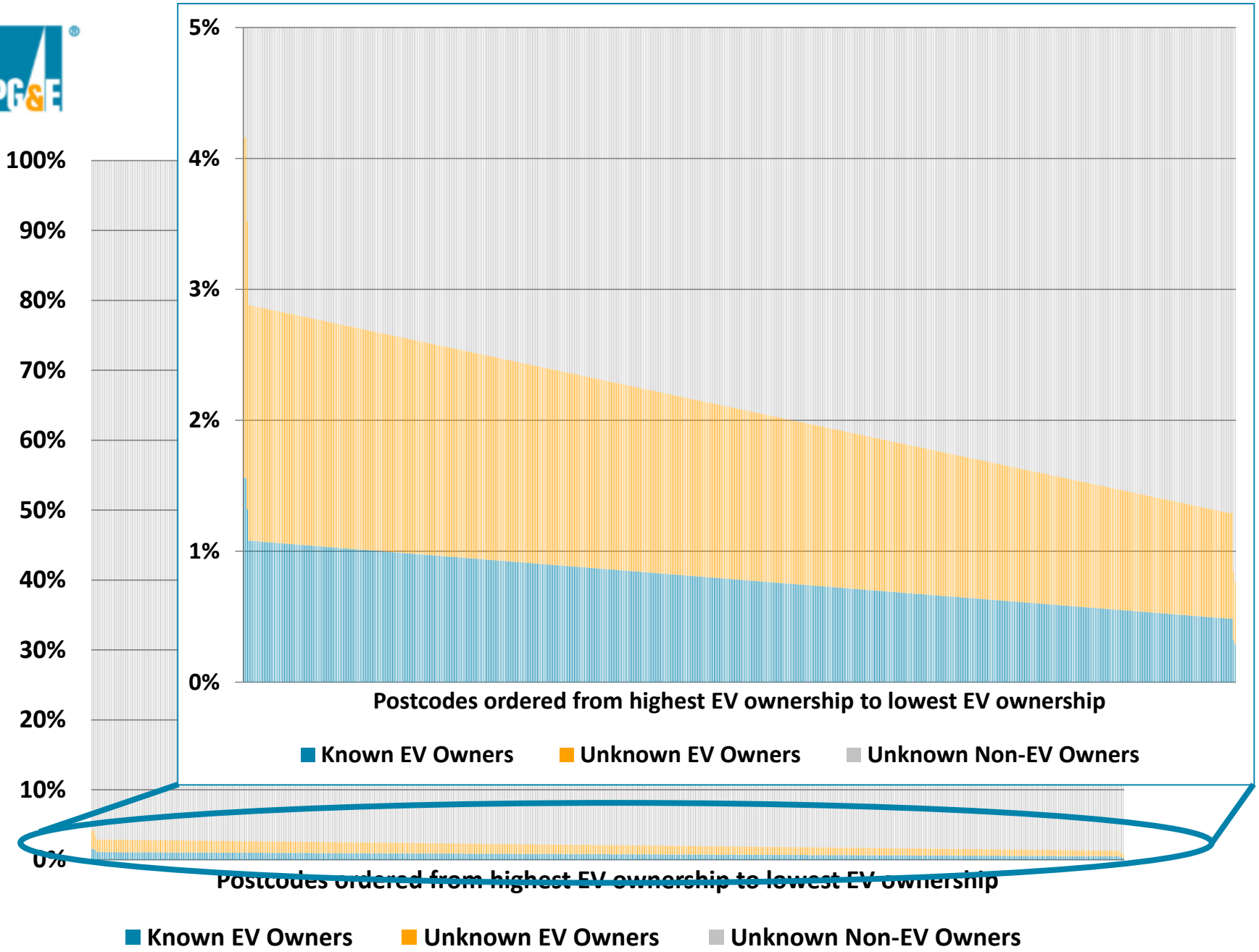
A representative sample of the PG&E population with known EV and non-EV customers does not exist

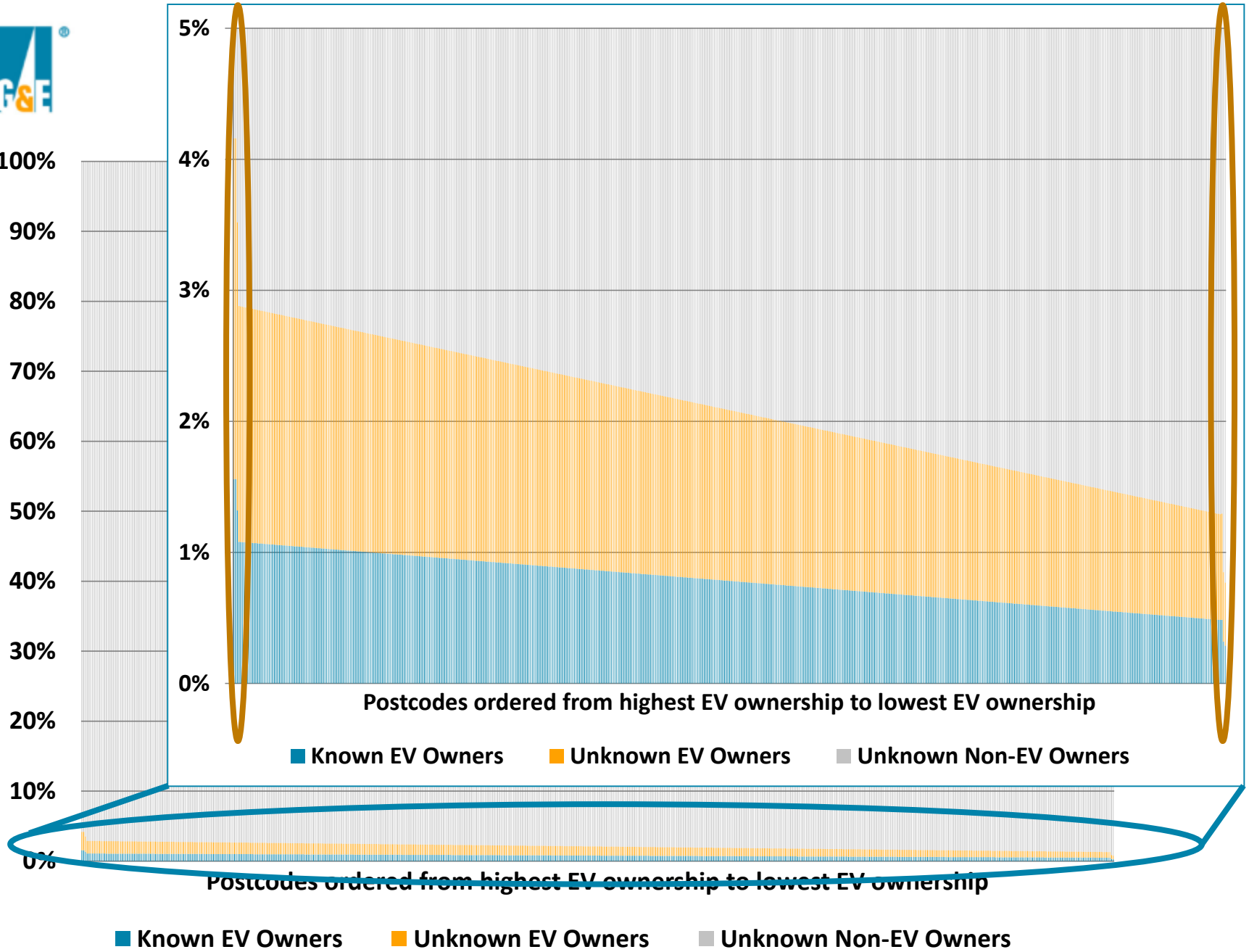
- Customers who purchase EV under no obligation to disclose purchase to utility or sign up for a TOU rate
- Only subset of EV population is known – estimate approximately 1/3 of EV population on EV TOU rate. . .*likely not representative of EV customers not on a TOU rate due to unobservable characteristics related to decision to enroll on EV TOU and effects of TOU rate on load*
- Even if EV TOU customers representative of all EV customers in general population, a sample of customers that definitely did not own EVs necessary to estimate any classification algorithm
- Remaining 2/3 EV owners unidentified in general residential population of approximately 6.4 million residential customers

- To address, construct 3 data sets:
  - EV rebate applications for 589 zip codes in PG&E territory
  - Customers on EV TOU rates

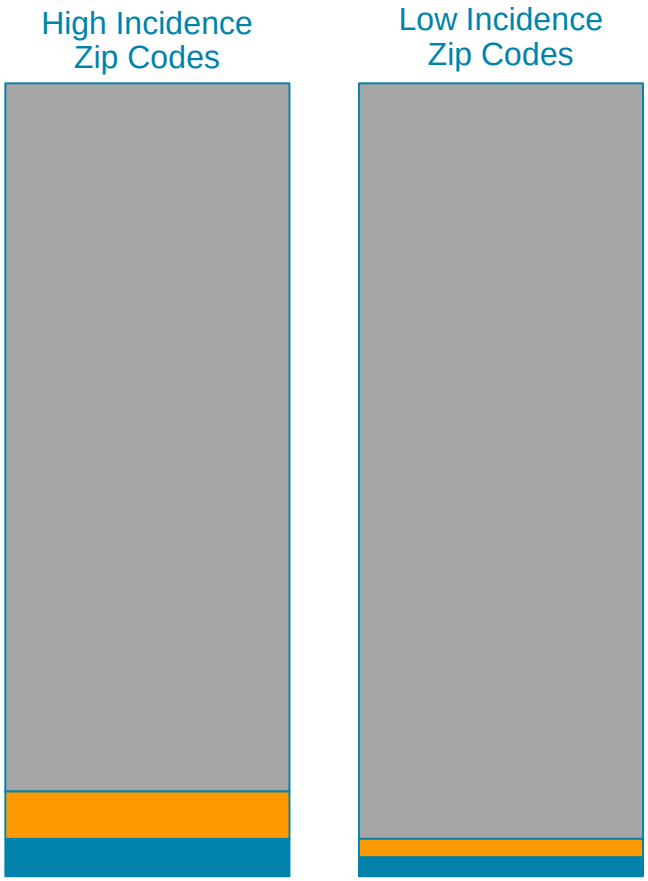
Data set #1


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Postcodes ordered from highest EV ownership to lowest EV ownership
Known EV Owners
Unknown EV Owners
Unknown Non-EV Owners


5%
4%
3%
2%
1%
0%
Postcodes ordered from highest EV ownership to lowest EV ownership
Known EV Owners
Unknown EV Owners
Unknown Non-EV Owners
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Postcodes ordered from highest EV ownership to lowest EV ownership
Known EV Owners
Unknown EV Owners
Unknown Non-EV Owners


5%
4%
3%
2%
1%
0%
Postcodes ordered from highest EV ownership to lowest EV ownership
Known EV Owners
Unknown EV Owners
Unknown Non-EV Owners
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Postcodes ordered from highest EV ownership to lowest EV ownership
Known EV Owners
Unknown EV Owners
Unknown Non-EV Owners

Data set #2

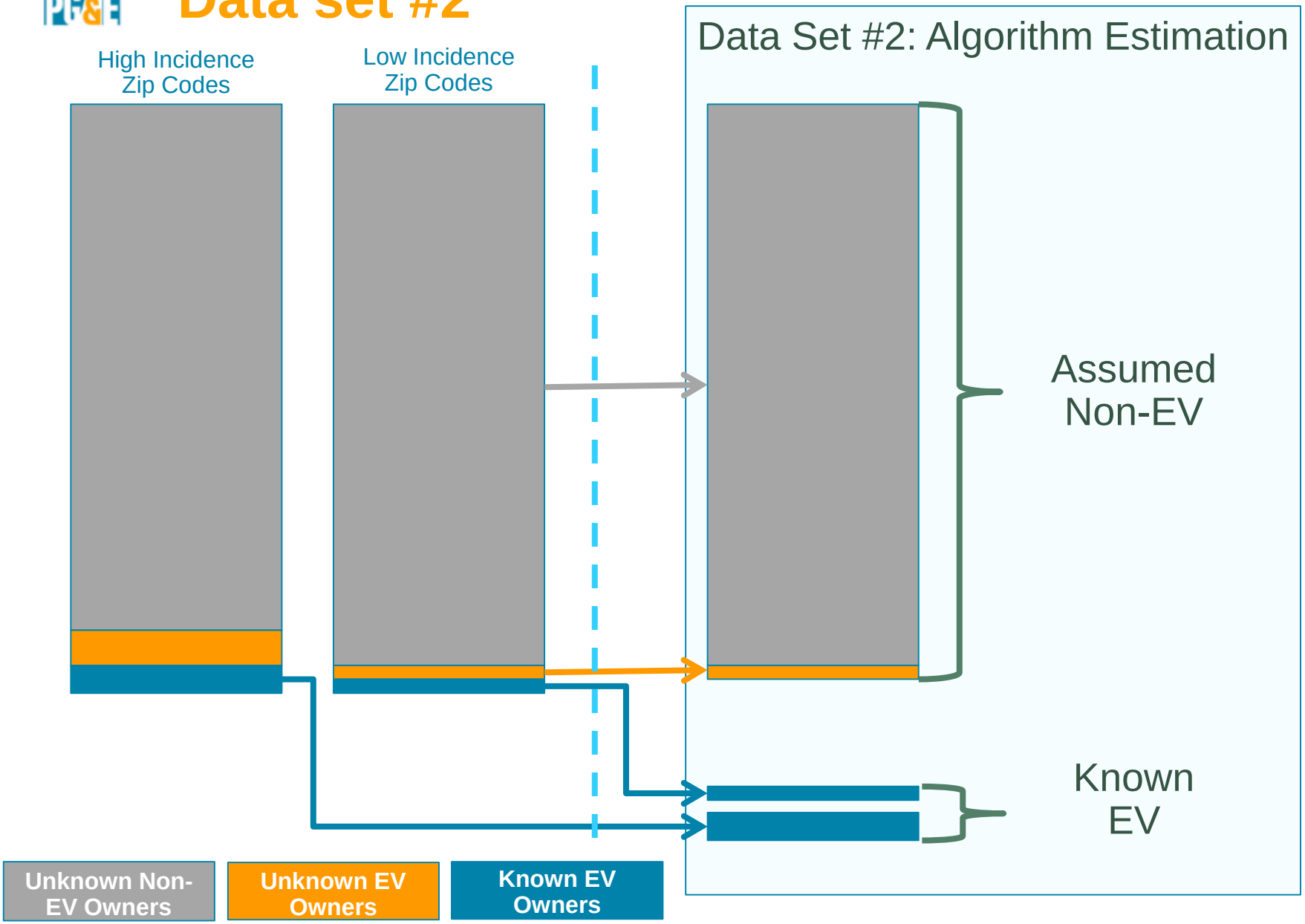


Unknown Non-EV Owners

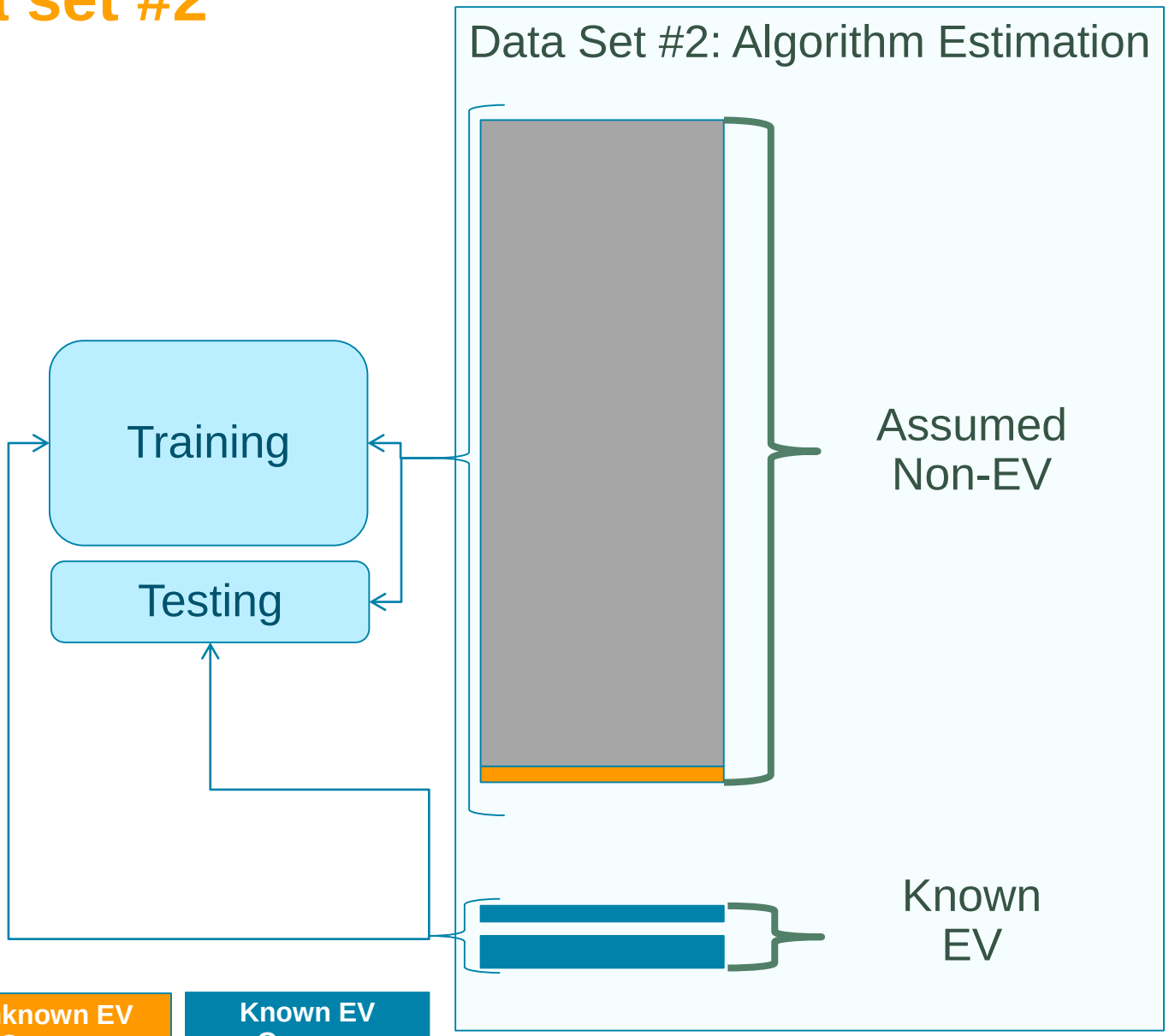
Unknown EV Owners

Known EV Owners

Data set #2


High Incidence Zip Codes
Low Incidence Zip Codes
Data Set #2: Algorithm Estimation
Assumed Non-EV
Known EV
Unknown Non-EV Owners
Unknown EV Owners
Known EV Owners

Data set #2

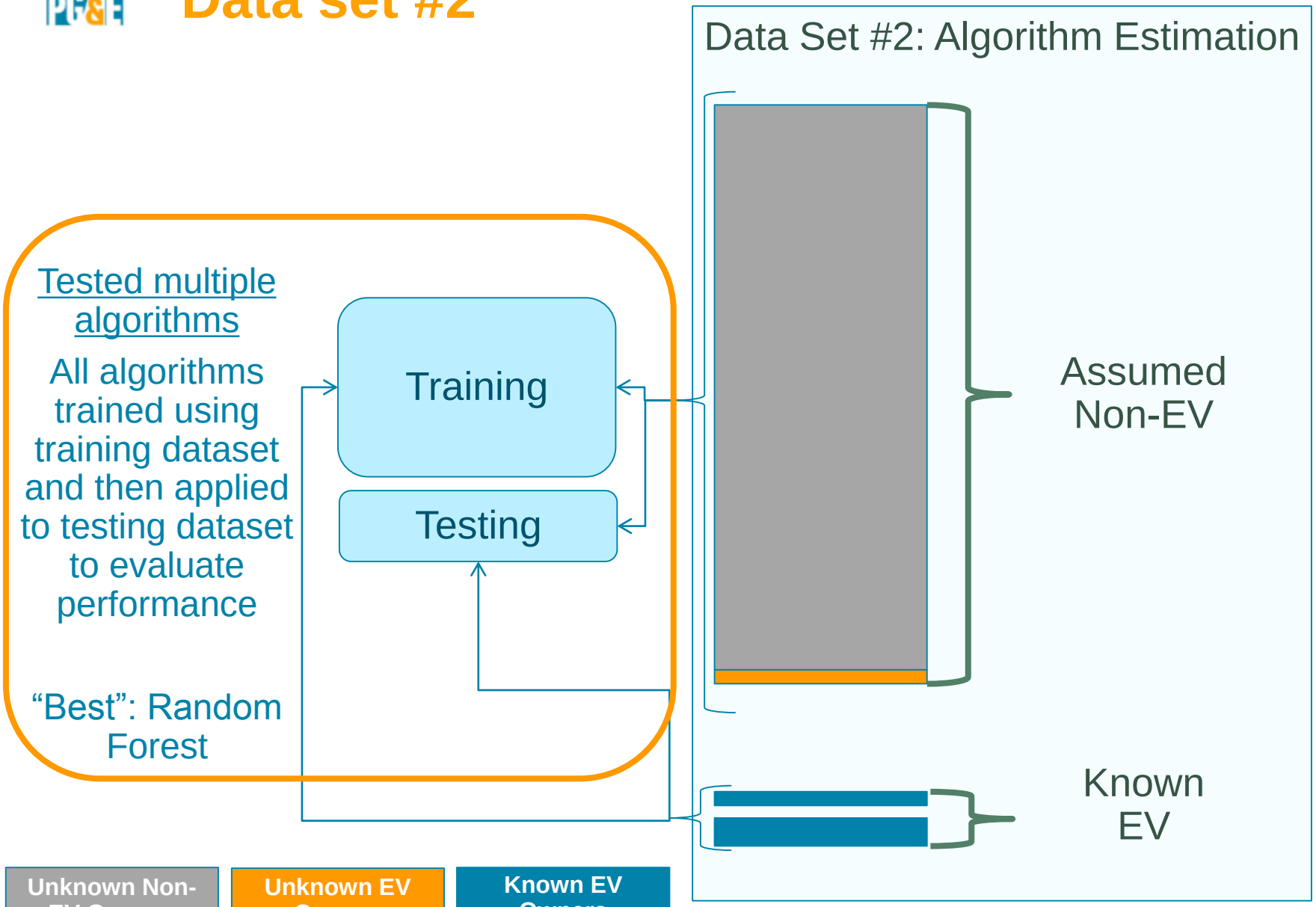


Unknown Non-EV Owners

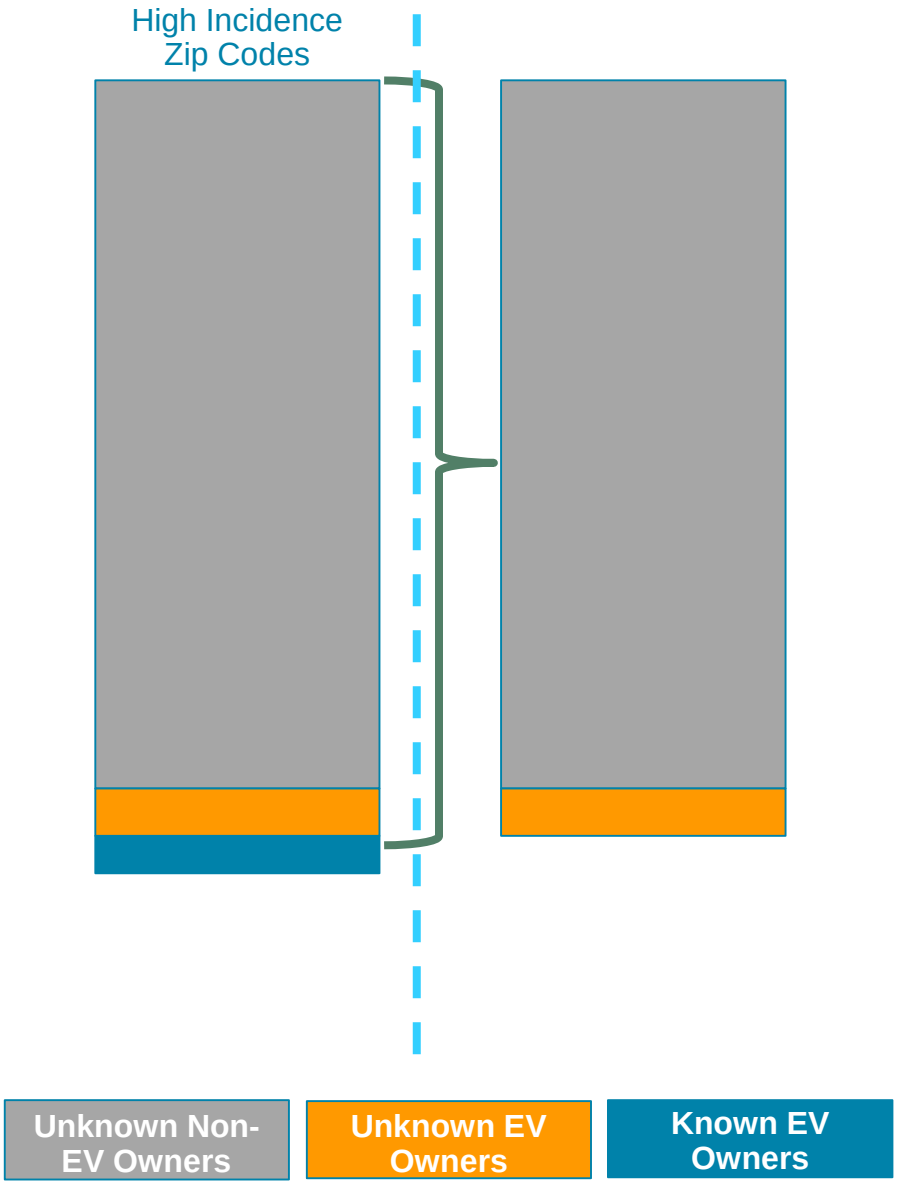
Unknown EV Owners

Known EV Owners

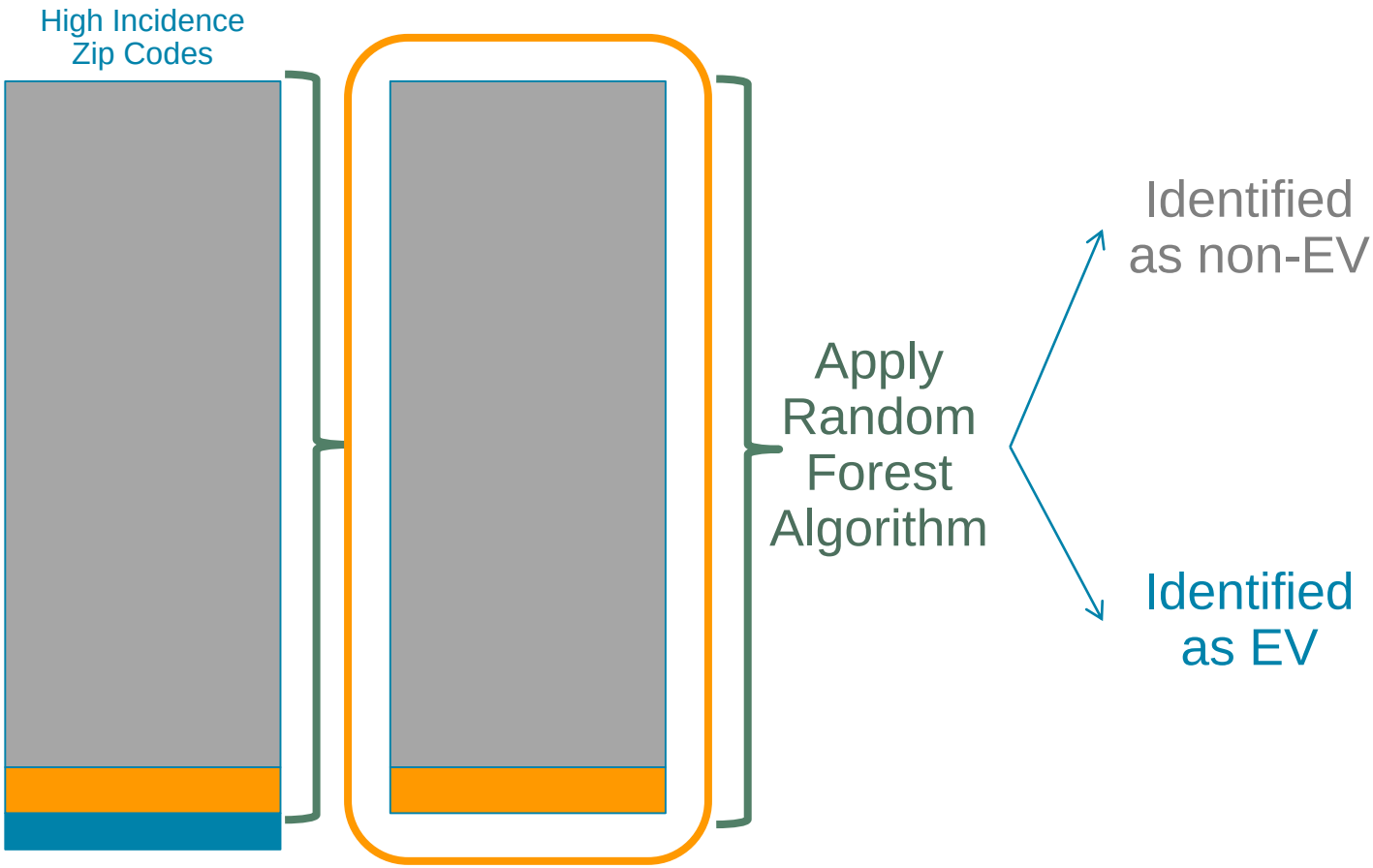
Data set #2


Data Set #2: Algorithm Estimation
Tested multiple algorithms
All algorithms trained using training dataset and then applied to testing dataset to evaluate performance
"Best": Random Forest
Training
Testing
Assumed Non-EV
Known EV
Unknown Non-EV Owners
Unknown EV Owners
Known EV Owners

Data set #3

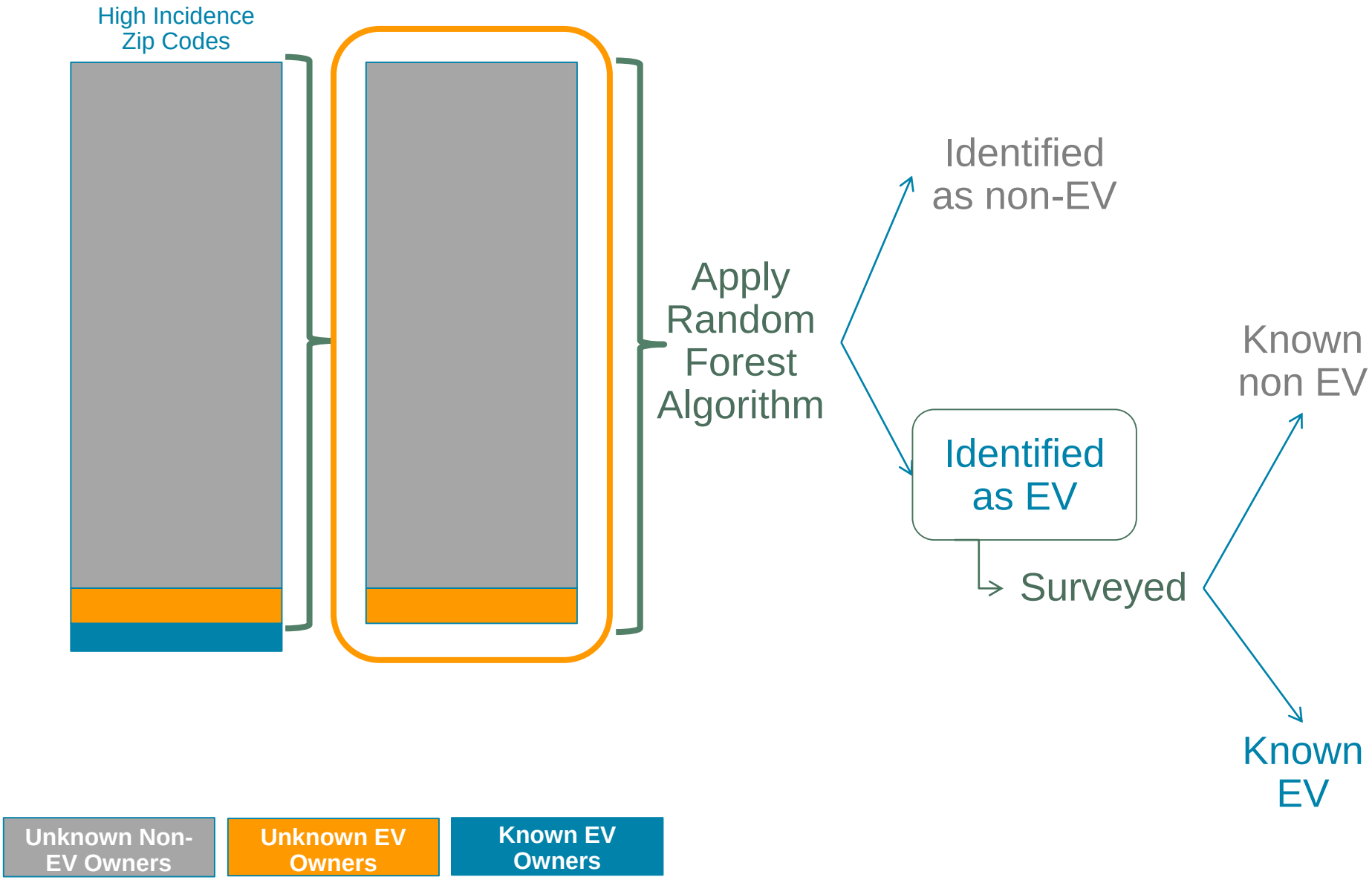

High Incidence
Zip Codes
Unknown Non-EV Owners
Unknown EV Owners
Known EV Owners

Data set #3



Unknown Non-EV Owners | Unknown EV Owners | Known EV Owners

Data set #3


High Incidence Zip Codes
Apply Random Forest Algorithm
Identified as non-EV
Identified as EV
Surveyed
Known non EV
Known EV
Unknown Non-EV Owners
Unknown EV Owners
Known EV Owners

Predictors of EV ownership

A wide variety of algorithms were explored and implemented

- Residential Class
- Percent of Charging Days
- Standard Deviation of Average Nighttime kW
- Standard Deviation of Normalized Average Nighttime kW
- kW1-kW24
- Average kW
- Average Day Time kW
- Average Night Time kW
- Average Maximum kW
- Average Night Time Maximum kW
- Normalized Average Night Time kW
- Normalized Average Night Time Maximum kW
- Standard Deviation of Normalized Average Nighttime Maximum kW
- Total kWh
- Maximum Percent Change in Monthly Consumption
- Maximum Percent Year on Year Change in Monthly Consumption
- Average Percent Year on Year Change in Monthly Consumption
- Maximum Percent Year on Year Change in Residual Monthly Consumption
- Average Percent Year on Year Change in Residual Monthly Consumption
- Total kWh for 6 Months and 12 Months
- Load Duration Curve Parameters (beta0, beta2)
- RSS of Load Duration Curve

Predictors of EV ownership

Four predictors played the biggest role in predicting EV ownership:

- **Residential Class:** Whether a customer's home is detached or has a shared wall/common area
- **Percent of "Charging Days":** The percentage of nights over three months that exceed some specific threshold
- **Standard Deviation of Average Nighttime kW:** The variability of night time loads
- **Standard Deviation of Normalized Average Nighttime kW:** The variability of normalized night time loads

Challenge of low incidence rates

When the incidence rate is lower than the false positive rate, false positive tests will be more probable than true positive tests

Hypothetical Example

Incidence rate of a super rare disease: 1 in 1,000,000, or 0.0001%

Positive test accuracy rate: 99%

False positive rate: 1%

False negative rate: 0%

Challenge of low incidence rates

When the incidence rate is lower than the false positive rate, false positive tests will be more probable than true positive tests

Hypothetical Example

Incidence rate of a super rare disease: 1 in 1,000,000, or 0.0001%

Positive test accuracy rate: 99%

False positive rate: 1%

False negative rate: 0%

	Has condition	Does not have condition	Total
Test positive	**1**	**10,000**	**10,001**
Test negative	0	989,999	989,999
Total	1	999,999	1,000,000

10,001 people test positive, but on 1 person has the condition.

Algorithm scenarios and results

- The percentage of households that have EVs is low: estimated to be 5% in zip codes with highest adoption rates.
- Given a 5% incidence rate, if you had a predictor that was 95% accurate, half of your results would still be false positives.

Algorithm scenarios and results

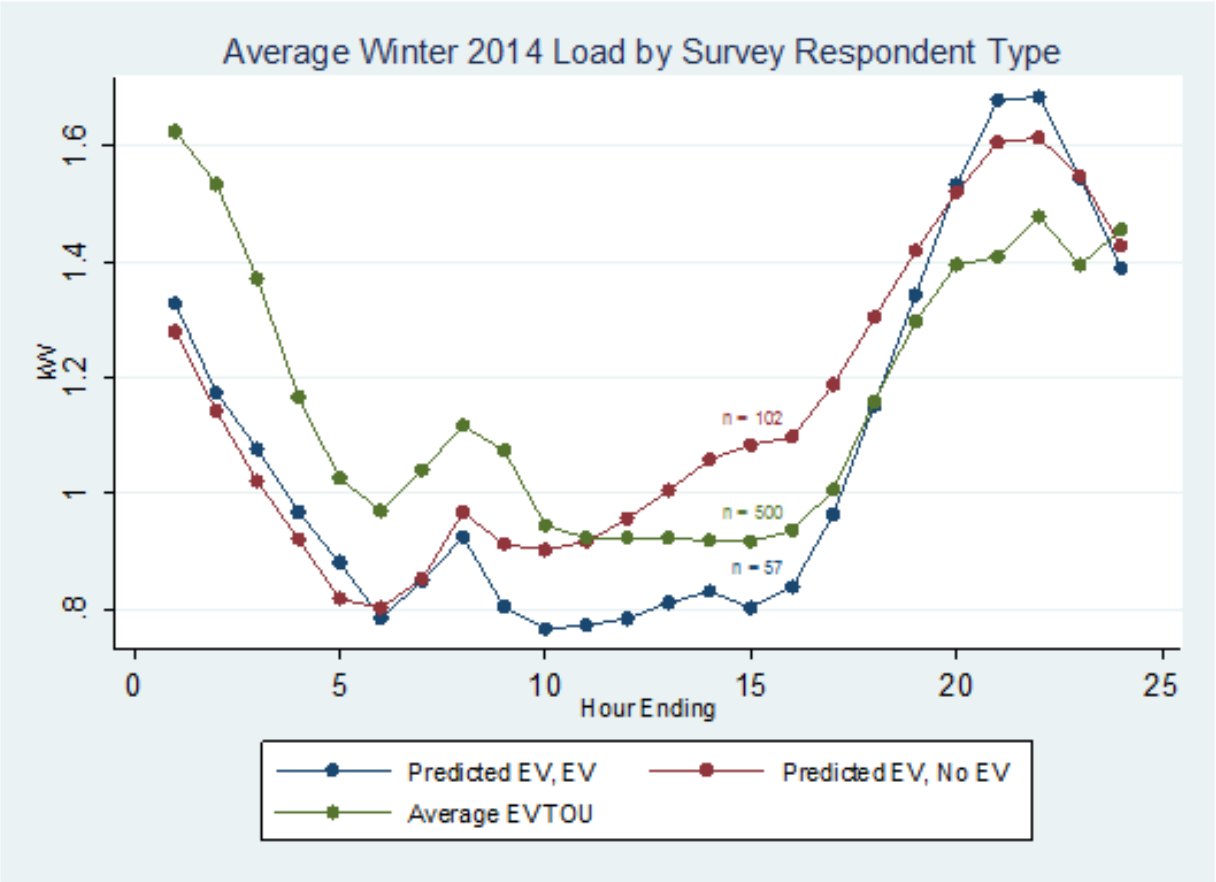
- The percentage of households that have EVs is low: estimated to be 5% in zip codes with highest adoption rates.
- Given a 5% incidence rate, if you had a predictor that was 95% accurate, half of your results would still be false positives.

	Perfect specificity scenario*	Preliminary algorithm results	Wave 1 and 2 results (n=852)	Wave 3 results (n= 439)
Population	150,000			
Incidence rate	2.2% (estimated)			
EV population	3,300 (estimated)			
% in survey sample who have EVs	100%	37.1%	30.2%	38.3%
Positive test accuracy rate	100%	96.6%	95.4% (estimated)	96.8% (estimated)

* False negative rate is fixed at 11% for all scenarios

Preliminary load comparison

The preliminary results indicate that EV owners on an EV TOU rate behave differently than EV owners on E-1. They tend to charge more during the off peak period and less in the afternoon and early evening.



Conclusions

- As expected, identifying EV owners is difficult in large part due to the low incidence rate
- That said, the initial results provide a benchmark to improve upon, and any predictive algorithm's success rate will improve as EV adoption grows
- Significantly more effective way to identify EV owners
- Initial results indicate that load patterns of EV owners on EV TOU and EV owners on E-1 differ significantly – EV owners on EV TOU responding to signal