

ACTIVE TRANSPORTATION COUNTS FROM EXISTING ON-STREET SIGNAL AND DETECTION INFRASTRUCTURE

NORTHWEST TRANSPORTATION CONFERENCE

MARCH 5, 2024

Research Team:

Sirisha Kothuri, PhD, Senior Research Associate, PI, PSU

Patrick Singleton, PhD, Associate Professor, Co-PI, USU

Mahyar Vahedi Saheli, Graduate Research Assistant, USU

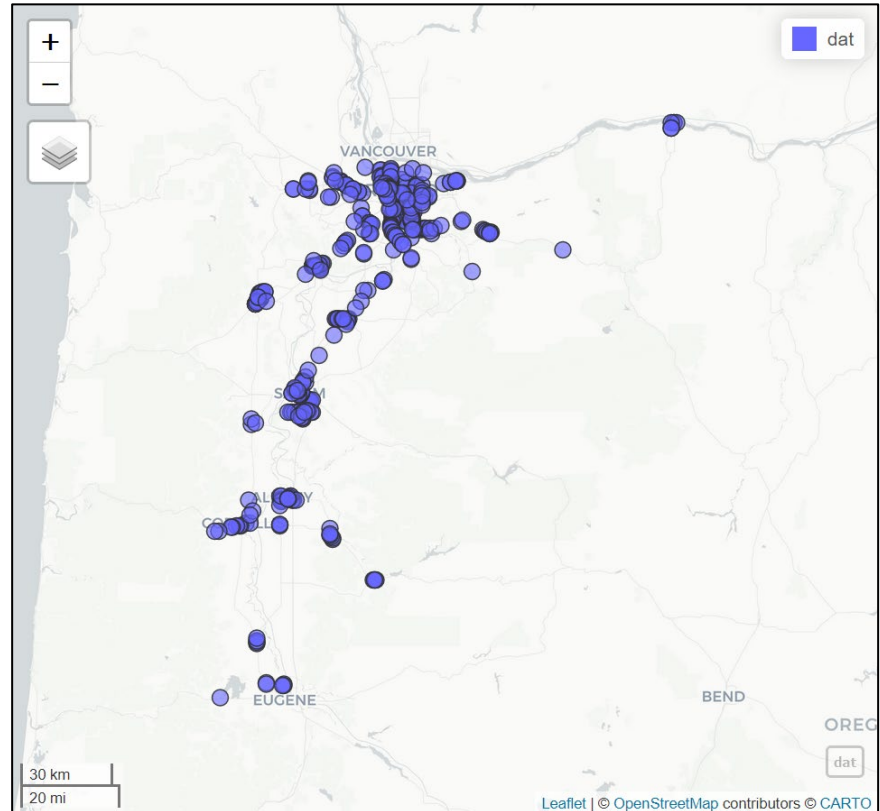
Elizabeth Yates, Graduate Research Assistant, PSU

Objectives

- Explore the feasibility of collecting pedestrian data from existing on-street infrastructure.
- Develop adjustment factors to convert pedestrian data to actual pedestrian counts.
- Determine the transferability of the methods developed and the efforts needed to apply these methods statewide.
- Develop a workflow to integrate pedestrian traffic counts into ODOT's enterprise traffic data system.

Inventory of Sites

- Agency contacts
 - ODOT
 - Portland (PBOT)
 - Corvallis
 - Eugene
 - Salem
 - Washington County
- 803 → 433 signals
 - Willamette Valley or close to Portland
 - High-resolution data
 - Some ped. activity



Site Selection Process

- ODOT, Portland (PBOT), & Washington County
- Stratified sampling by
 - Place Type →
 - Low / non-MPO
 - Medium MPO
 - High MPO
 - Pedestrian activity
 - Low: <250
 - Medium: 250-1000
 - High: >1000

#	Place Type	Group
1	Rural	Low / non-MPO
2	Isolated City	
3	Rural near Major Center	
4	City near Major Center	
5&6	MPO Low Density	
7	MPO Residential	Medium MPO
8	MPO Employment	
9	MPO Mixed Use	High MPO
10	MPO TOD	

<https://www.oregon.gov/lcd/CL/Pages/Place-Types.aspx>

Study Locations

Potential sample (45-55)

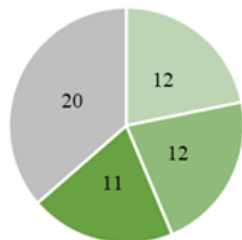
	Low (10-100)	Medium (100-250)	High (250+)	NA
Low or Medium Non-MPO	4	4	3	2
Medium MPO	4	4	4	7
High	4	4	4	11

Final sample (49)

	Low (10-100)	Medium (100-250)	High (250+)	NA
Low or Medium Non-MPO	4	4	3	1
Medium MPO	5	4	4	3
High	4	4	3	10

Pedestrian activity

- Low (10-100)
- Medium (100-250)
- High (250+)
- NA



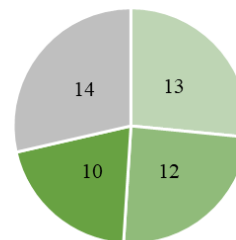
Place type

- Low or Medium Non-MPO
- Medium MPO
- High



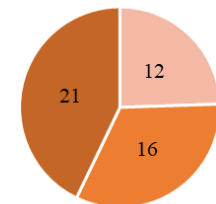
Pedestrian activity

- Low (10-100)
- Medium (100-250)
- High (250+)
- NA

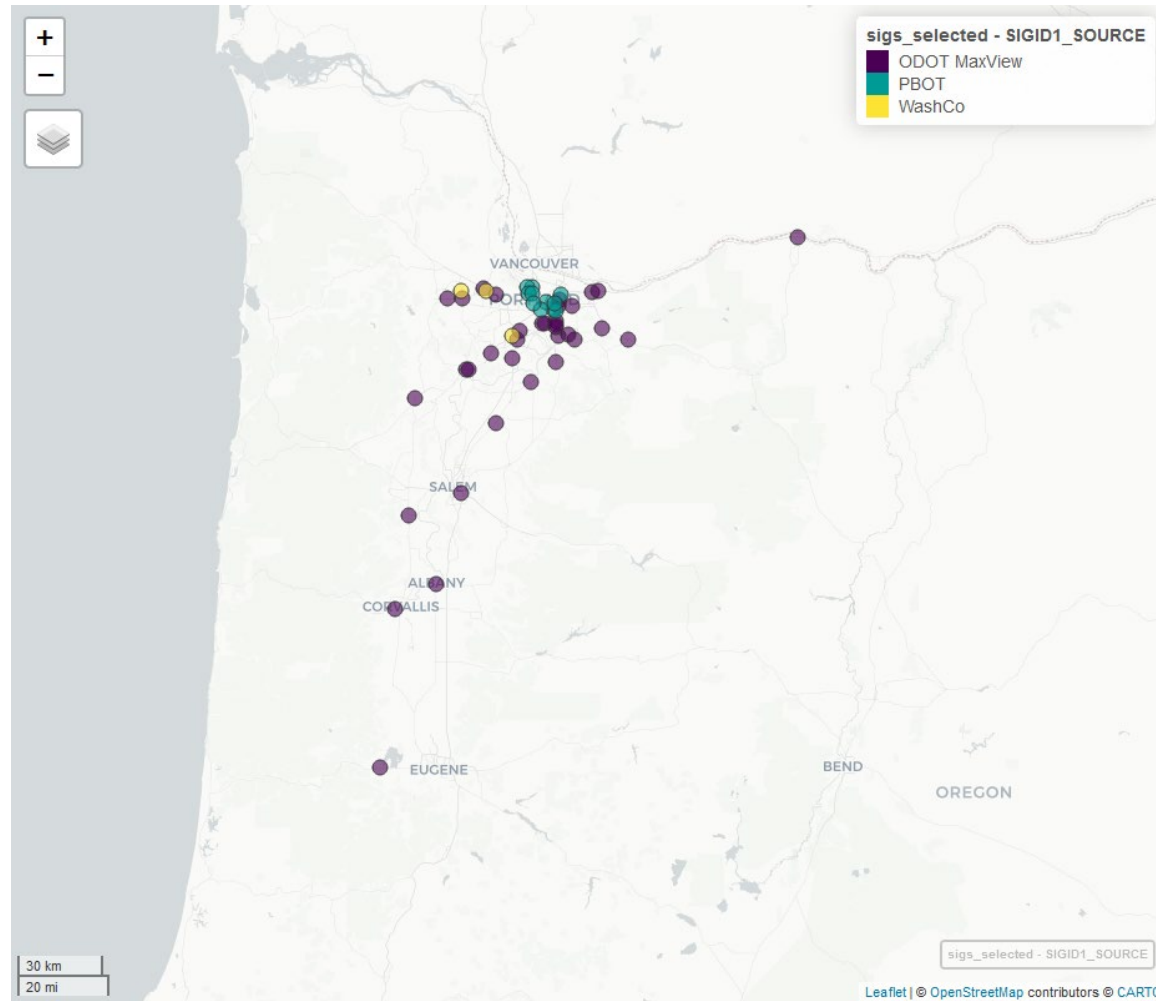


Place type

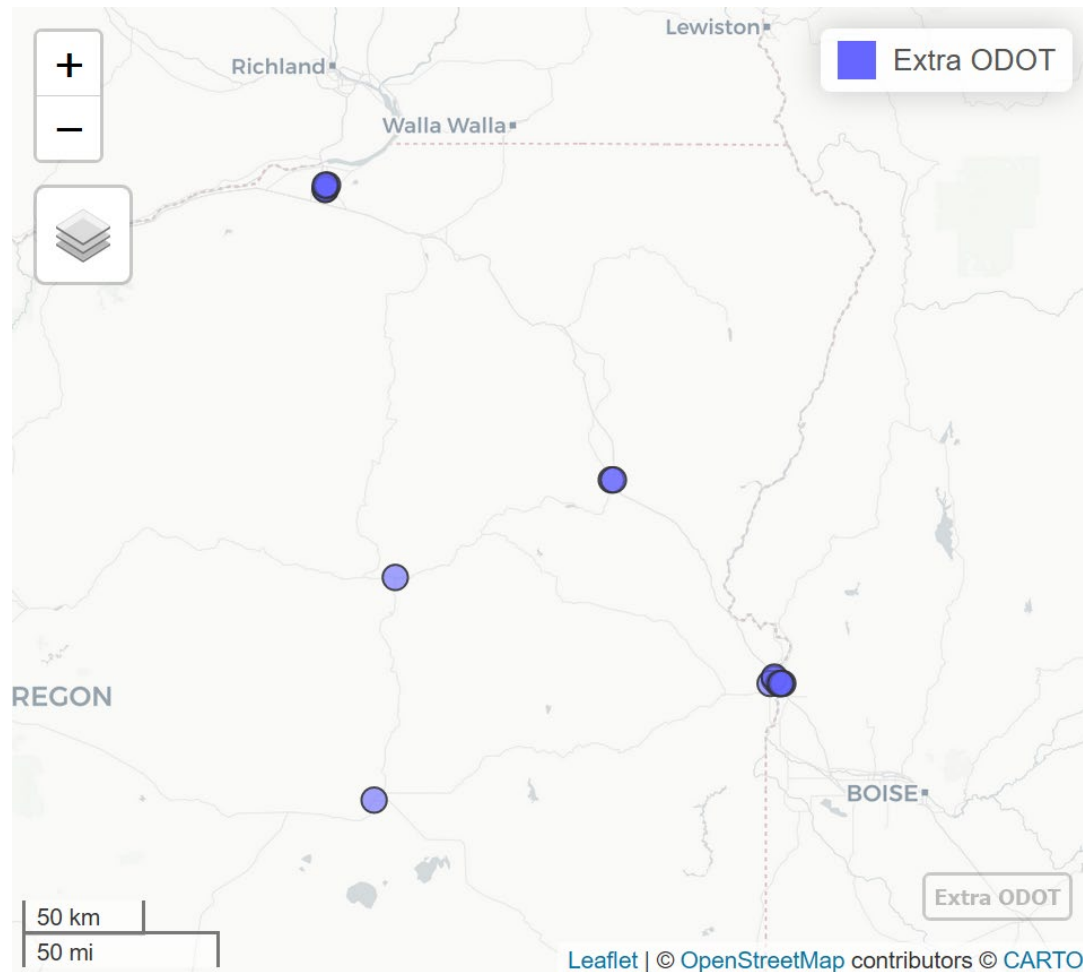
- Low or Medium Non-MPO
- Medium MPO
- High



Selected Sites



Extra Sites



Data Extraction Process

100440 - Ellsworth at 6th-Data Extraction Form.xlsm - Excel

File Home Insert Page Layout Formulas Data Review View Developer Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells

N1

1	Intersection ID	100440
2	Intersection Name	Ellsworth at 6th
3	Crosswalk Name	
4	Direction of travel	
5	Crosswalk user type	Pedestrian
6	Number of users	1
7	Date	2022-06-22
8	Timestamp 1	
9	Timestamp 2	
10	Notes	
11	Submit	

2022-06-07 9:00:04 PM

Form Stored data

Ready Accessibility: Investigate

Data Extraction Process

	A	B
1	Intersection ID	100440
2	Intersection Name	Ellsworth at 6th
3	Crosswalk Name	
4	Direction of travel	North South East West
5	Crosswalk user type	
6	Number of users	1
7	Date	2022-06-22
8	Timestamp 1	
9	Timestamp 2	
10	Notes	
11-14	Submit	

	A	B
1	Intersection ID	100440
2	Intersection Name	Ellsworth at 6th
3	Crosswalk Name	
4	Direction of travel	
5	Crosswalk user type	S->N S<-N E->W E<-W
6	Number of users	
7	Date	2022-06-22
8	Timestamp 1	
9	Timestamp 2	
10	Notes	
11-14	Submit	

	A	B
1	Intersection ID	100440
2	Intersection Name	Ellsworth at 6th
3	Crosswalk Name	
4	Direction of travel	
5	Crosswalk user type	Pedestrian Bicycle Wheelchair Skateboard Scooter Other
6	Number of users	
7	Date	
8	Timestamp 1	
9	Timestamp 2	
10	Notes	
11-14	Submit	

	A	B
1	Intersection ID	100440
2	Intersection Name	Ellsworth at 6th
3	Crosswalk Name	
4	Direction of travel	
5	Crosswalk user type	Pedestrian
6	Number of users	1
7	Date	2022-06-22
8	Timestamp 1	
9	Timestamp 2	
10	Notes	
11-14	Submit	

Data Extraction Process



Validation Process

	B	C	D	E	F
1	Intersection ID	Intersection Name	Crosswalk Name	Direction of travel	Crosswalk user type
2	100412	OR224 at Springwater	East	N-->S	Pedestrian
3	100412	OR224 at Springwater	South	E-->W	Bicycle
4	100412	OR224 at Springwater	East	N-->S	Bicycle
5	100412	OR224 at Springwater	East	N<--S	Bicycle
6	100412	OR224 at Springwater	East	N-->S	Pedestrian
7	100412	OR224 at Springwater	South	E<--W	Pedestrian
8	100412	OR224 at Springwater	South	E-->W	Pedestrian
9	100412	OR224 at Springwater	South	E<--W	Pedestrian
10	100412	OR224 at Springwater	South	E-->W	Pedestrian
11	100412	OR224 at Springwater	East	N<--S	Pedestrian
12	100412	OR224 at Springwater	East	N-->S	Pedestrian
13	100412	OR224 at Springwater	East	N<--S	Pedestrian
14	100412	OR224 at Springwater	South	E<--W	Pedestrian
15	100412	OR224 at Springwater	South	E<--W	Pedestrian
16	100412	OR224 at Springwater	South	E-->W	Pedestrian
17	100412	OR224 at Springwater	South	E-->W	Pedestrian
18	100412	OR224 at Springwater	East	N-->S	Pedestrian
19	100412	OR224 at Springwater	South	E<--W	Pedestrian
20	100412	OR224 at Springwater	East	N<--S	Pedestrian
21	100412	OR224 at Springwater	South	E<--W	Pedestrian
22	100412	OR224 at Springwater			Pedestrian
23	100412	OR224 at Springwater	South	E<--W	Pedestrian
24	100412	OR224 at Springwater	East	N-->S	Pedestrian
25	100412	OR224 at Springwater	East	N<--S	Pedestrian
26	100412	OR224 at Springwater	South	E-->W	Pedestrian
27	100412	OR224 at Springwater	East	N-->S	Pedestrian
28	100412	OR224 at Springwater	South	E-->W	Pedestrian
29	100412	OR224 at Springwater	East	N<--S	Pedestrian
30	100412	OR224 at Springwater	East	N<--S	Pedestrian
31	100412	OR224 at Springwater	South	E<--W	Pedestrian
32	100412	OR224 at Springwater	South	E-->W	Pedestrian

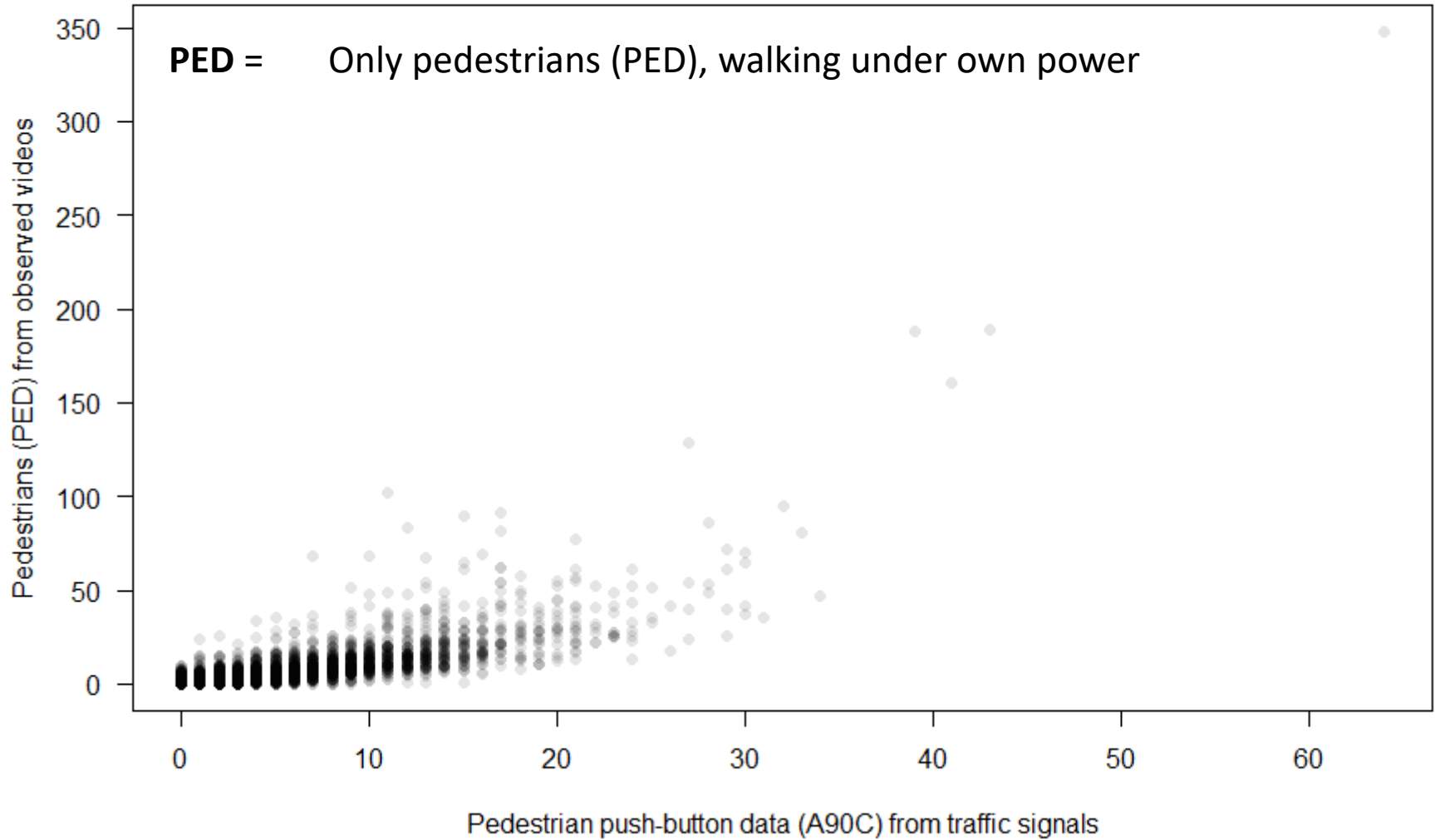
- As soon as the data extraction is finished for each intersection, **5% (at least 10 records)** of the data is **double-checked** by someone other than the person collected the data.
- If **systematic** problems are observed, **further validation** was carried out.

Descriptive statistics

- Total people counted
 - 35,767 pedestrians
 - 5,032 people bicycling
 - 449 people on scooters
 - 278 people on skateboards
 - 233 people in wheelchairs
 - 84 other crosswalk users (e.g., OneWheel)
- People counted, by intersection
 - Minimum = 5 (signal 100770, OR 201 & Washington, Ontario)
 - Maximum = 2,649 (signal 201026, Going & Interstate, Portland)
 - Mean = 658
 - Median = 388

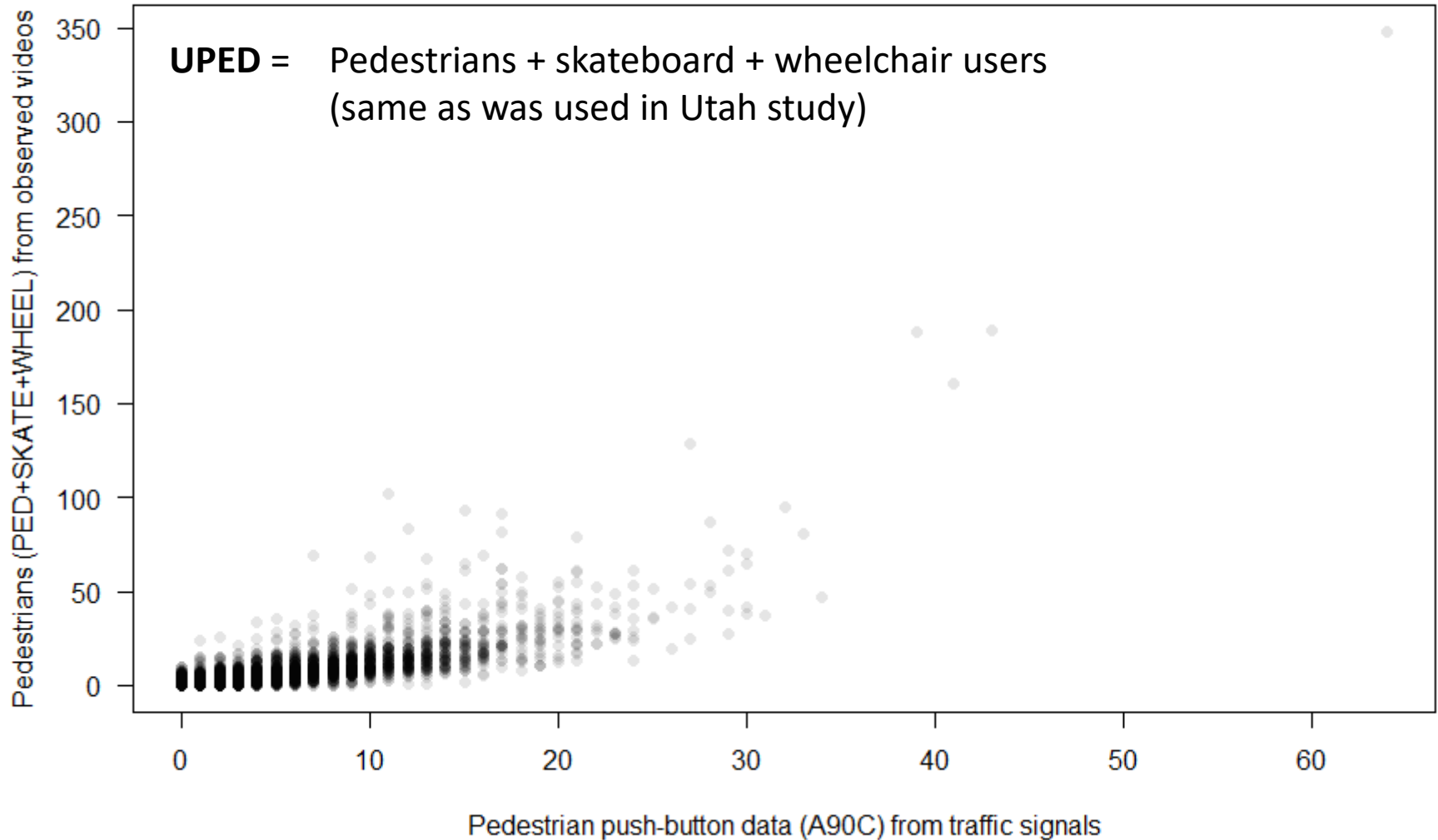
41,843 users

Pedestrian signal data collection in Oregon (final, 2023-05-09)



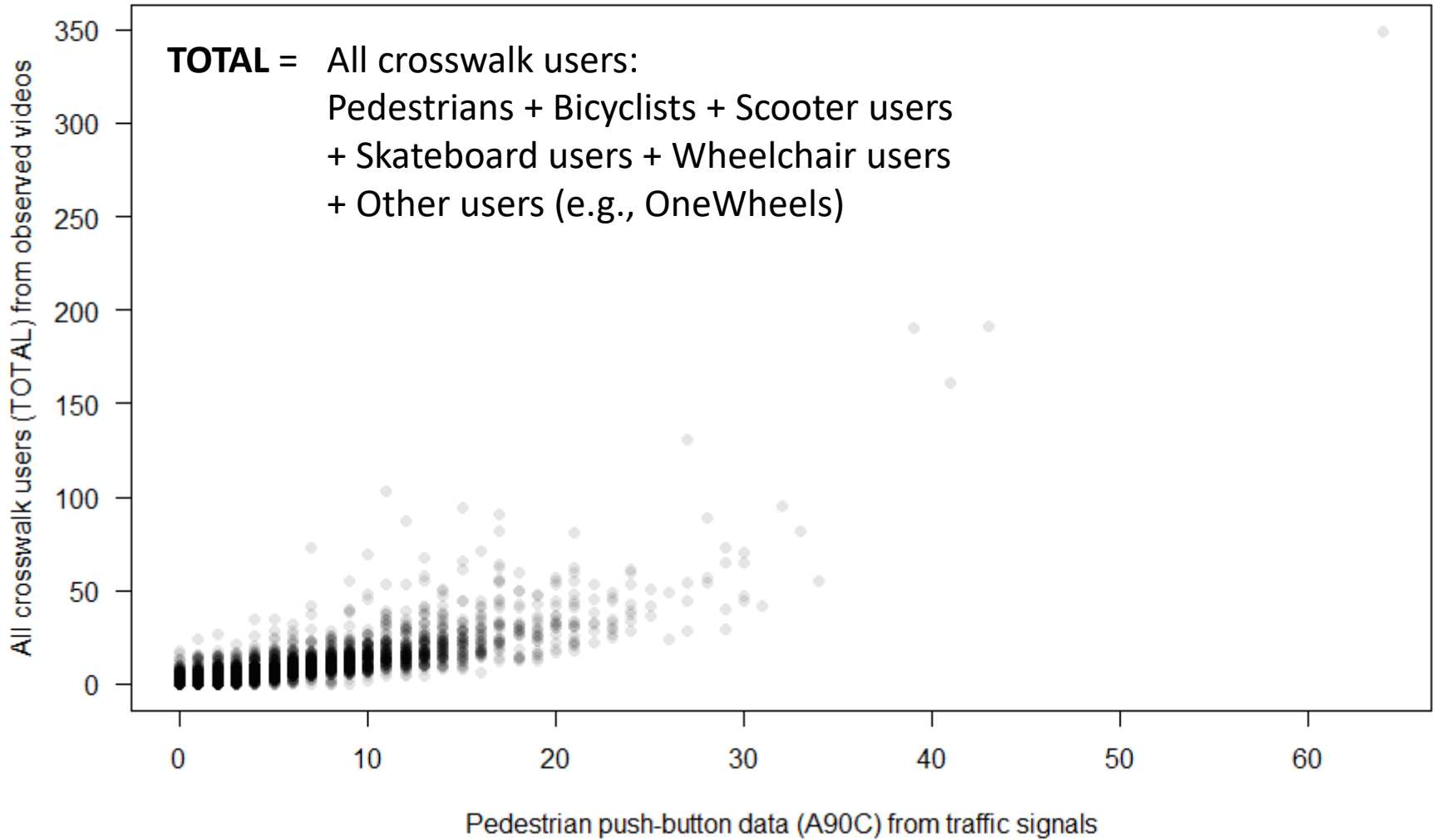
N = 8,620. Each observation represents one hour at one crosswalk at one signal.

Pedestrian signal data collection in Oregon (final, 2023-05-09)



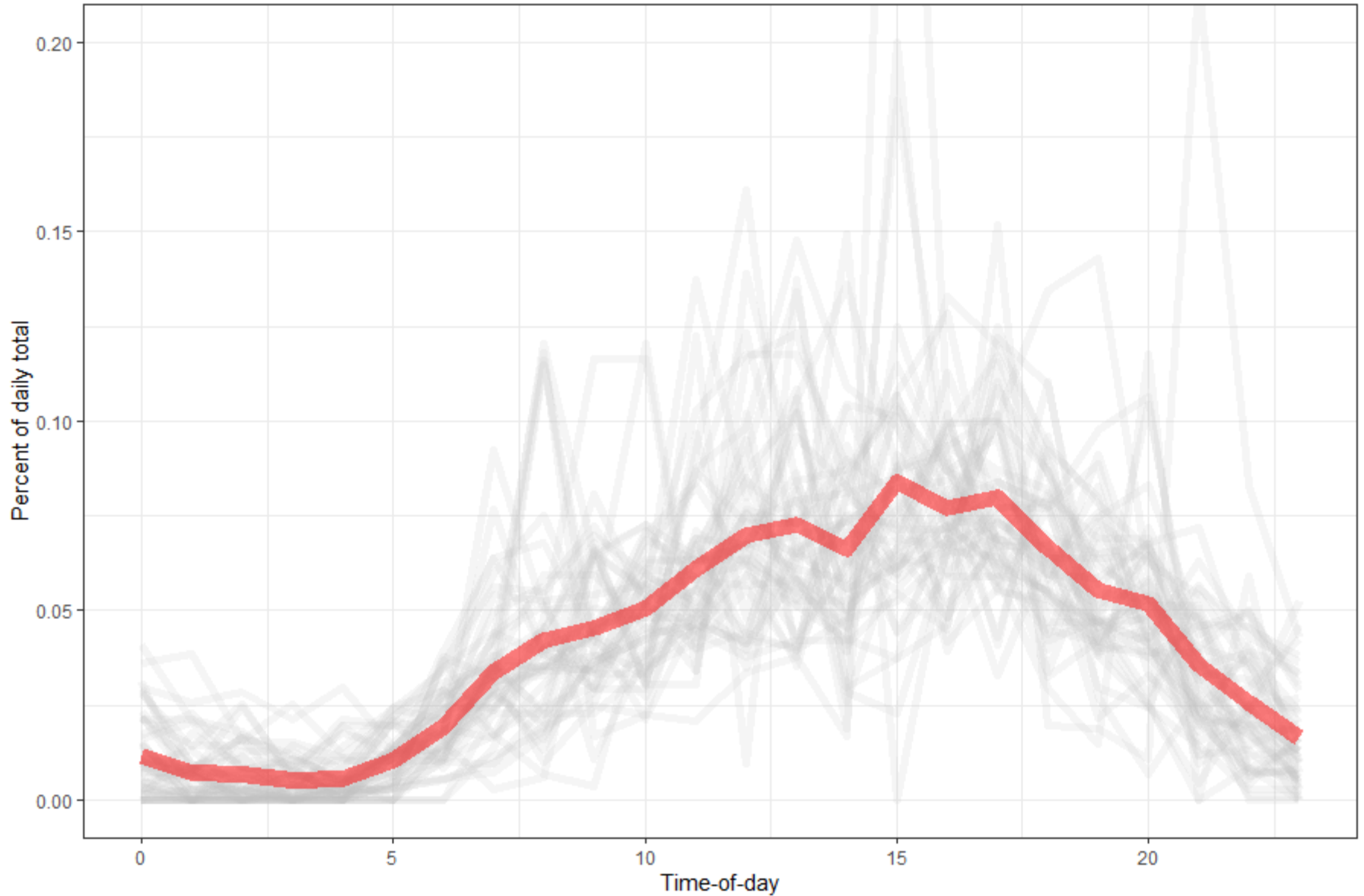
N = 8,620. Each observation represents one hour at one crosswalk at one signal.

Pedestrian signal data collection in Oregon (final, 2023-05-09)

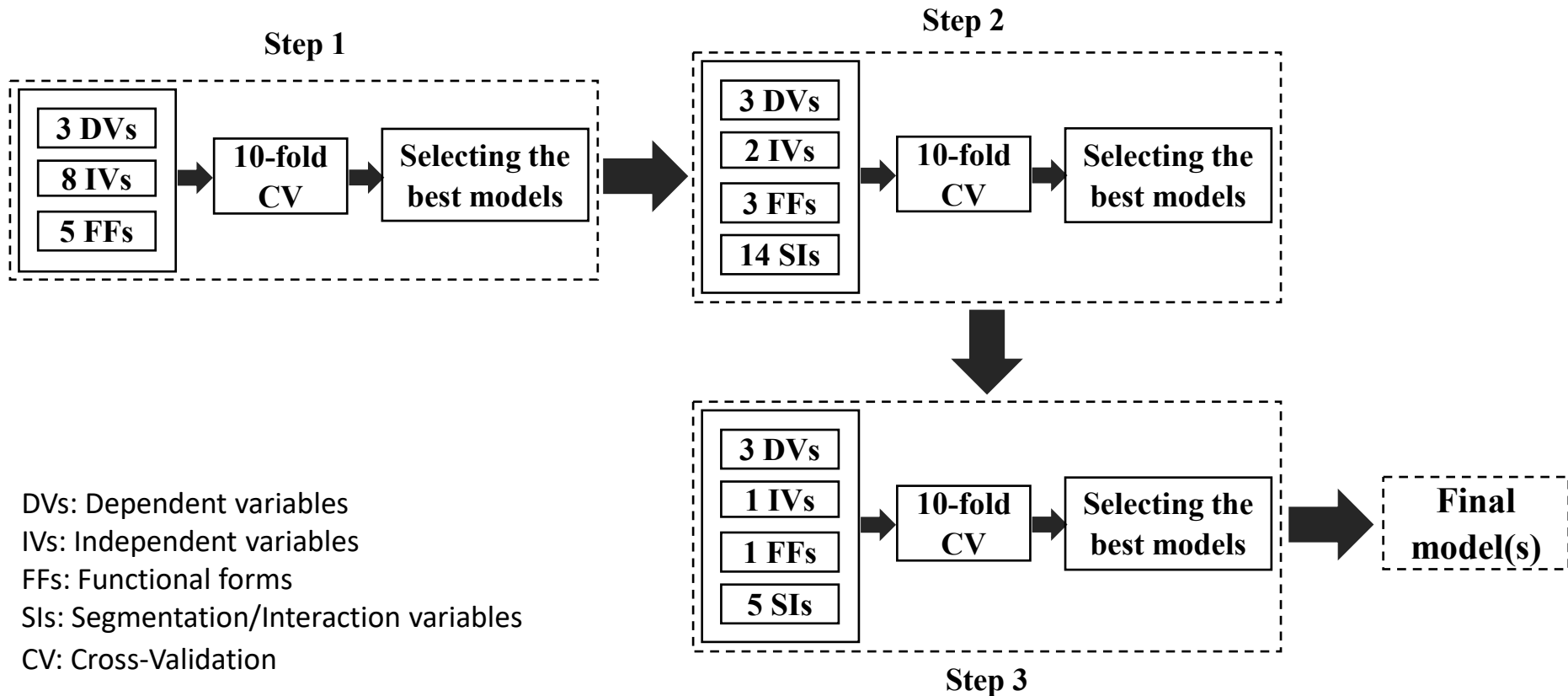


N = 8,620. Each observation represents one hour at one crosswalk at one signal.

Time-of-day distribution(s), by hour-of-day



Sequential Search Process



DVs: Dependent variables
IVs: Independent variables
FFs: Functional forms
SIs: Segmentation/Interaction variables
CV: Cross-Validation

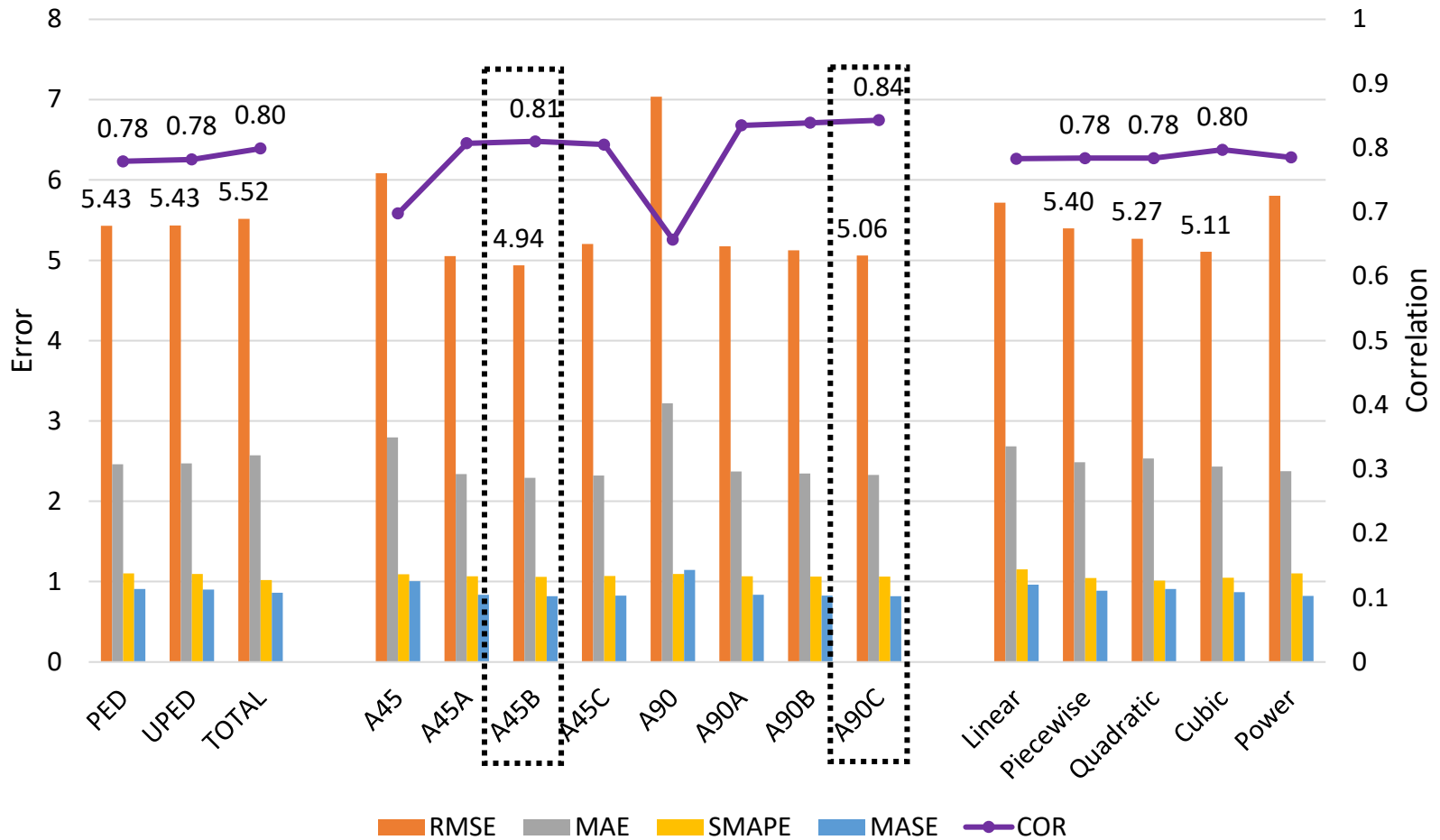
*** Please refer to previous slides for the full list of the DVs, IVs, FFs, and SIs.

Step 1: DVs, IVs, FFs

- 3 Dependent variables (DVs)
 - **PED, UPED*, TOTAL**
- 8 Independent variables (IVs)
 - **A45, A45A, A45B*, A45C, A90, A90A, A90B, A90C***
- 5 Functional forms (FFs)
 - **Linear, Piecewise*, Quadratic*, Cubic, Power**

Results for Step 1

Validation statistics using the hold-out validation data*



* Higher is better for "COR", and lower is better for the rest of the statistics

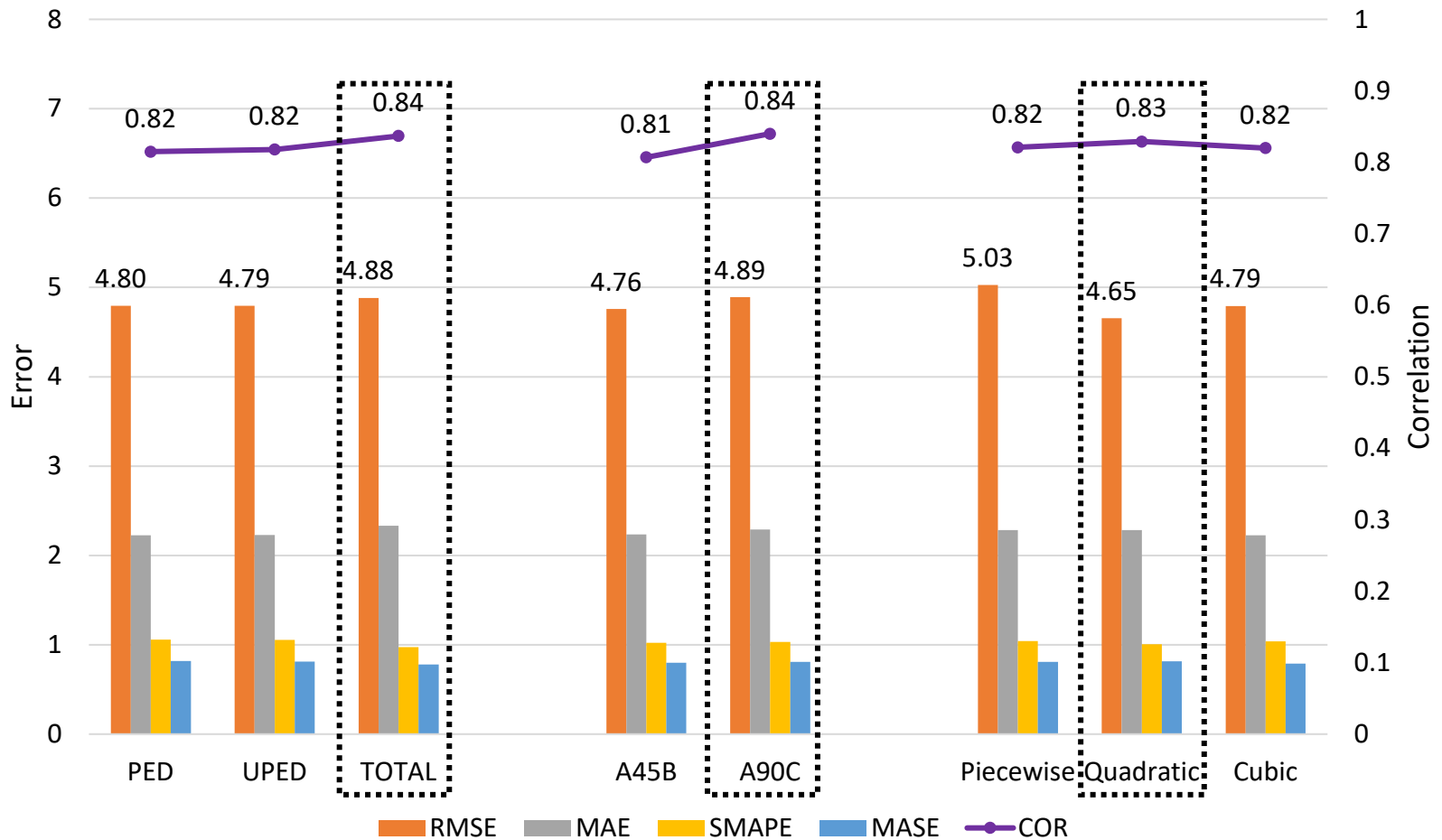
Step 2: DVs, IVs, FFs, and SIs

- 3 Dependent variables (DVs)
 - **PED, UPED*, TOTAL**
- 2 Independent variables (IVs)
 - **A45B*, A90C***
- 3 Functional forms (FFs)
 - **Piecewise*, Quadratic*, Cubic**
- 14 Segmentation/interaction variables (SIs)
 - **RECALL, CYCLE090, CYCLE120, PEAKAM, PEAKPM, PEAKAMP, OVERNIGHT, WEEKEND, ODOTREG, PLACETYPE3, TRANSIT4C, TRANSIT8C, EDTOTAL4C, EDTOTAL8C**

* Also used in Utah study.

Results for Step 2

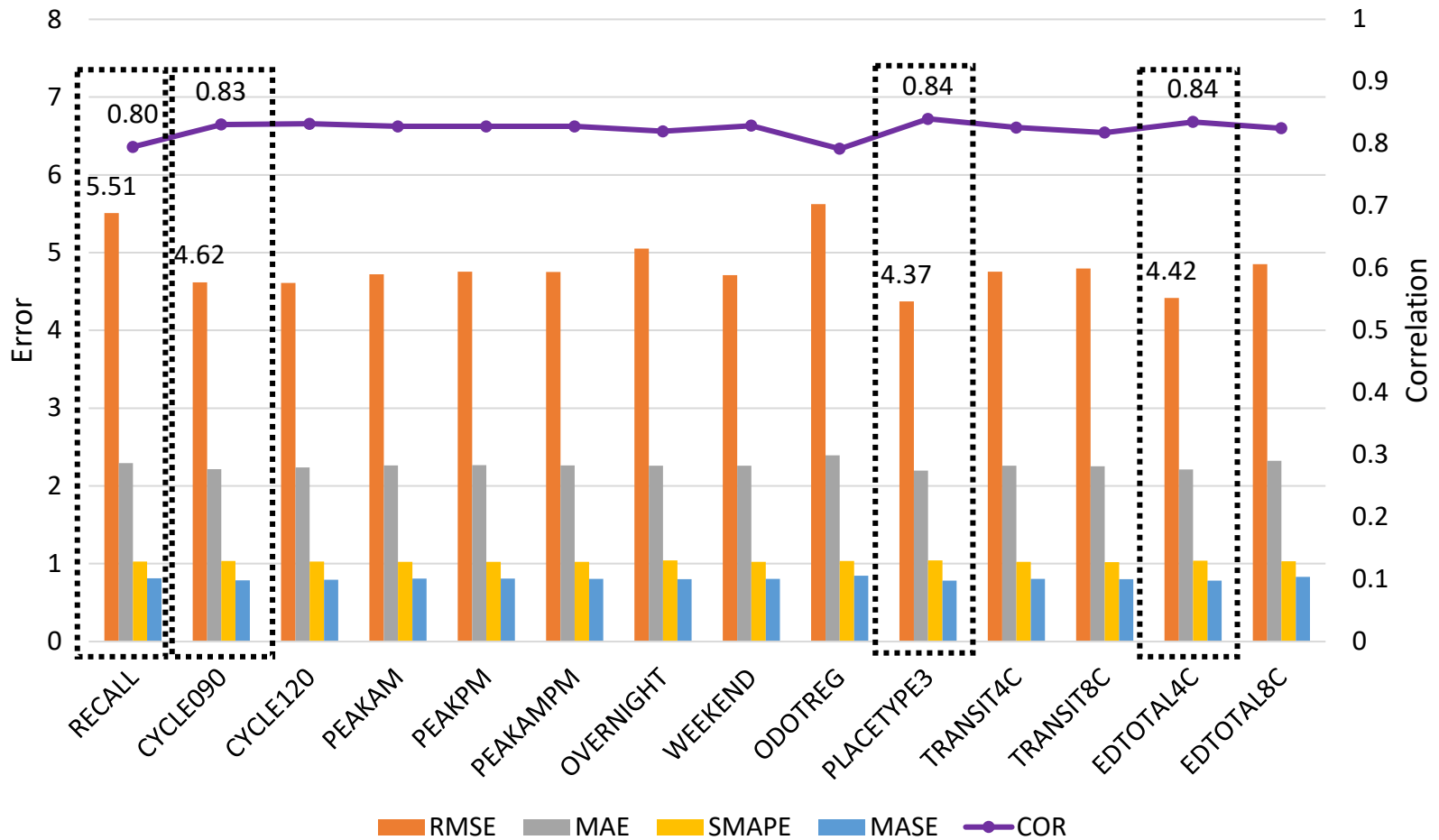
Validation statistics using the hold-out validation data*



* Higher is better for "COR", and lower is better for the rest of the statistics

Results for Step 2

Validation statistics using the hold-out validation data*



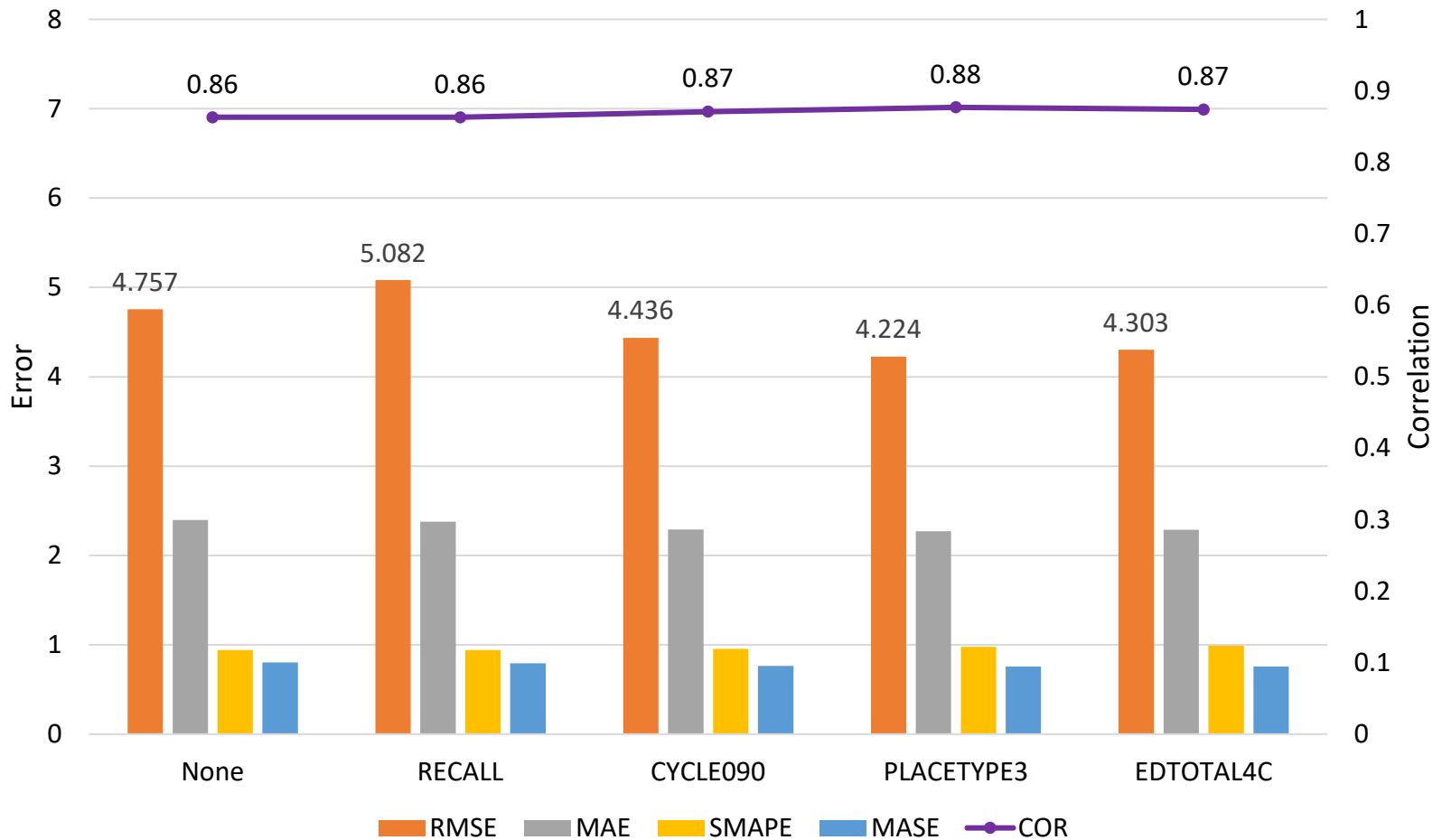
* Higher is better for "COR", and lower is better for the rest of the statistics

Step 3: DVs, IVs, FFs, and SIs

- 3 Dependent variables (DVs)
 - **PED, UPED*, TOTAL**
- 1 Independent variables (IVs)
 - **A90C***
- 1 Functional forms (FFs)
 - **Quadratic***
- 5 Segmentation/interaction variables (SIs)
 - **RECALL, CYCLE090, PLACETYPE3, EDTOTAL4C, No SI**

Results for Step 3

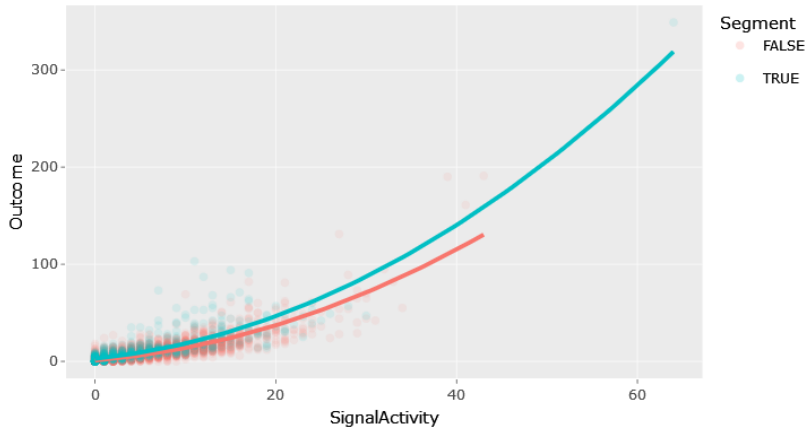
Validation statistics using the hold-out validation data*



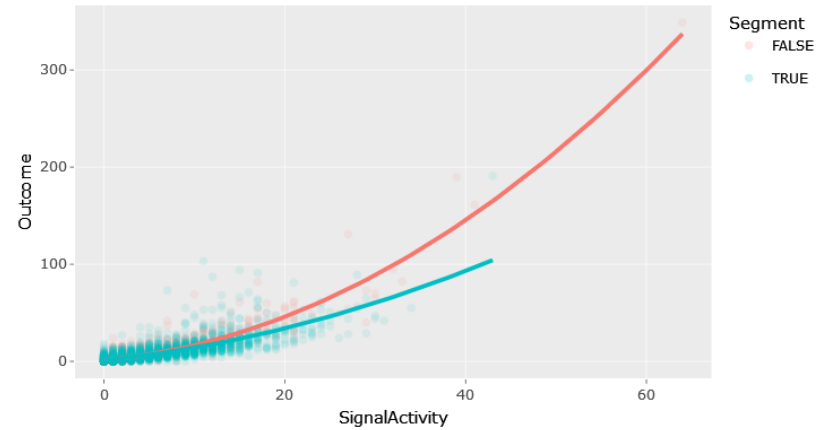
* Higher is better for "COR", and lower is better for the rest of the statistics

Results for Step 3

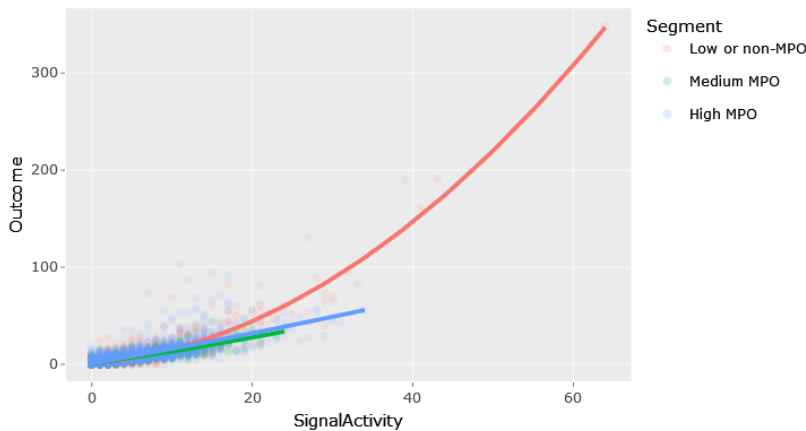
RECALL



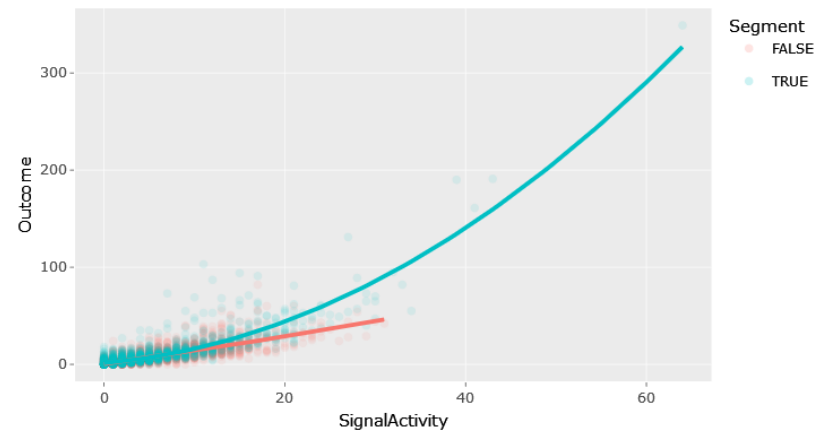
CYCLE090



PLACETYPE3



EDTOTAL4C



Results for Step 3

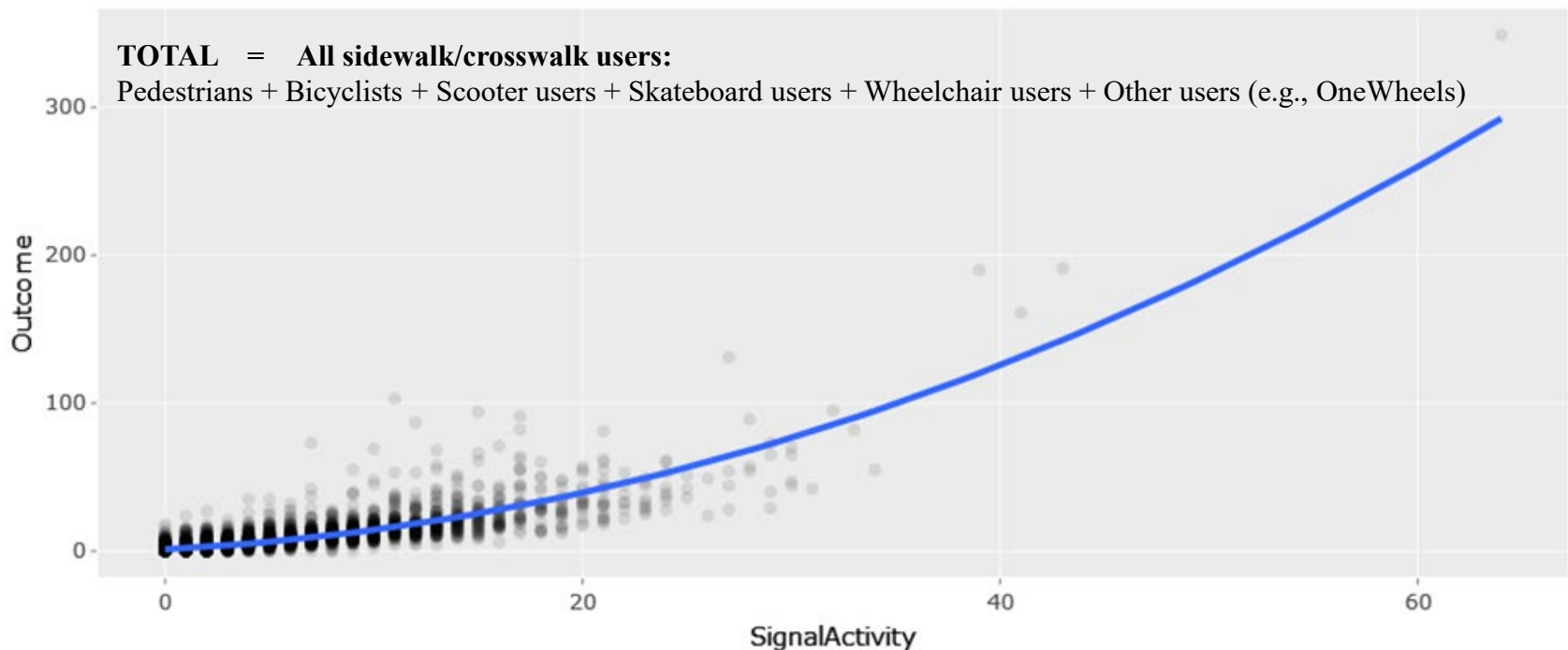
- Why not to move forward with SIs?
 - No large difference in model's predictions
 - Simplicity and interpretability of the results
 - Lack of longevity and transferability
 - No guarantee that these factors will remain the same in future years

Final results: recommended models

- 3 Dependent variables (DVs)
 - **PED, UPED*, TOTAL**
- 1 Independent variables (IVs)
 - **A90C***
- 1 Functional forms (FFs)
 - **Quadratic***
- No Segmentation/interaction variables (SIs)

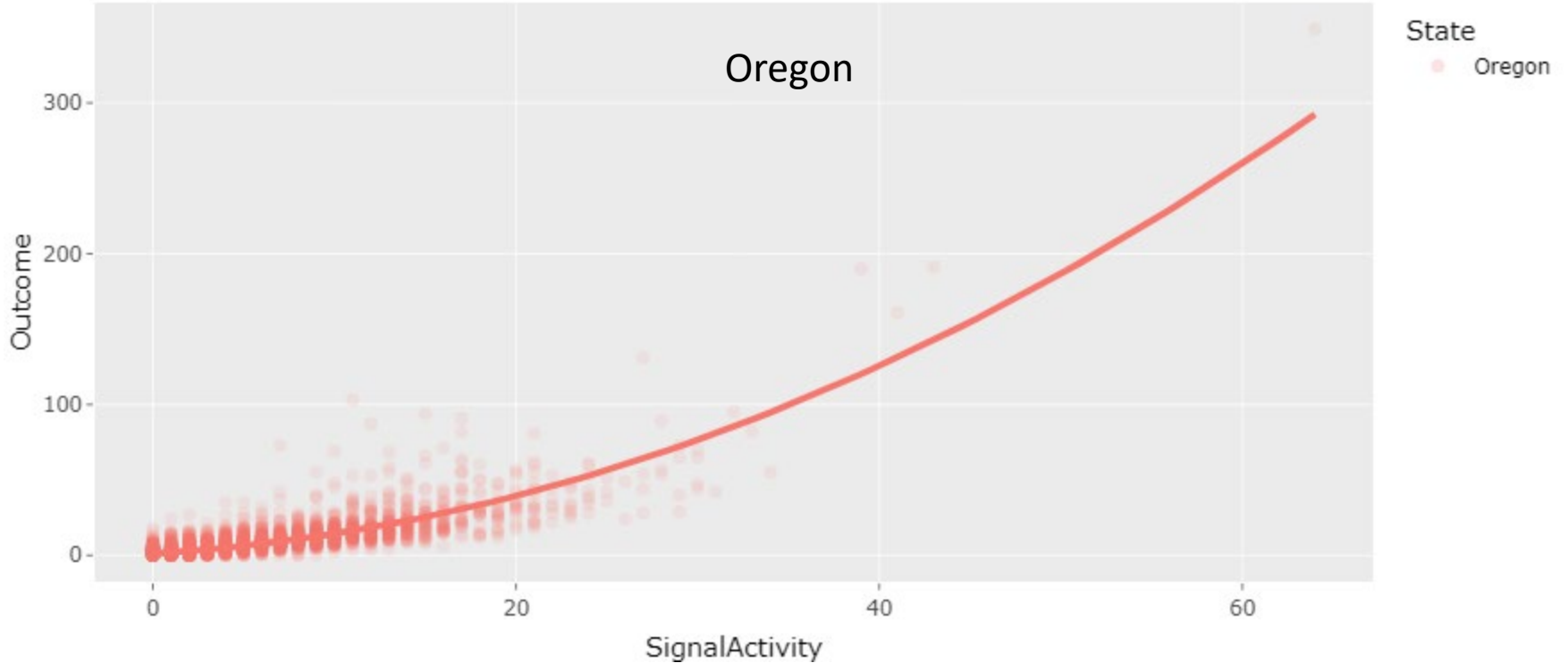
Results for Step 3

DV	N	R ²	COR	RMSE	MAE	SMAPE	MASE
TOTAL	8,546	0.7580	0.8628	4.7566	2.3990	0.9412	0.8027
UPED	8,546	0.7308	0.8339	4.7628	2.3328	1.0165	0.8517
PED	8,546	0.7278	0.8311	4.7697	2.3288	1.0229	0.8602



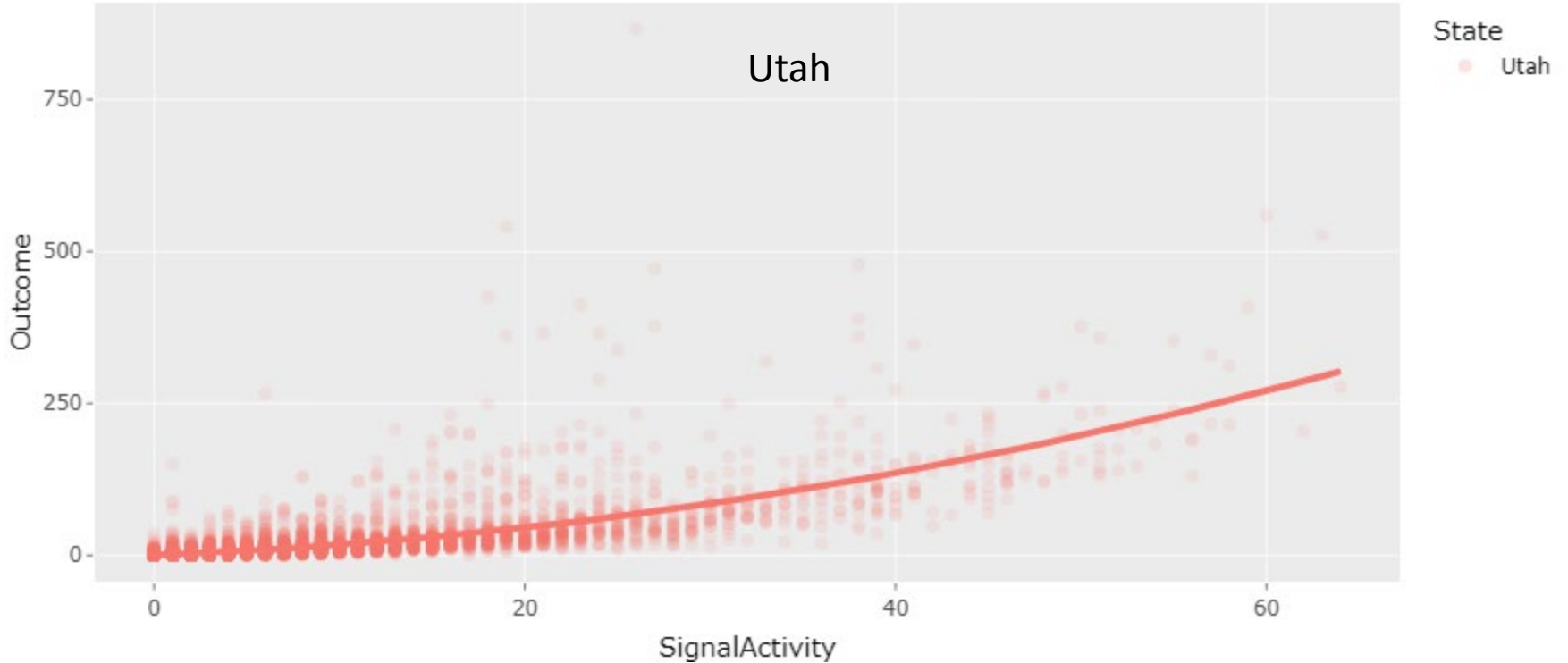
OR vs. UT

- How does Oregon data compare to Utah data?



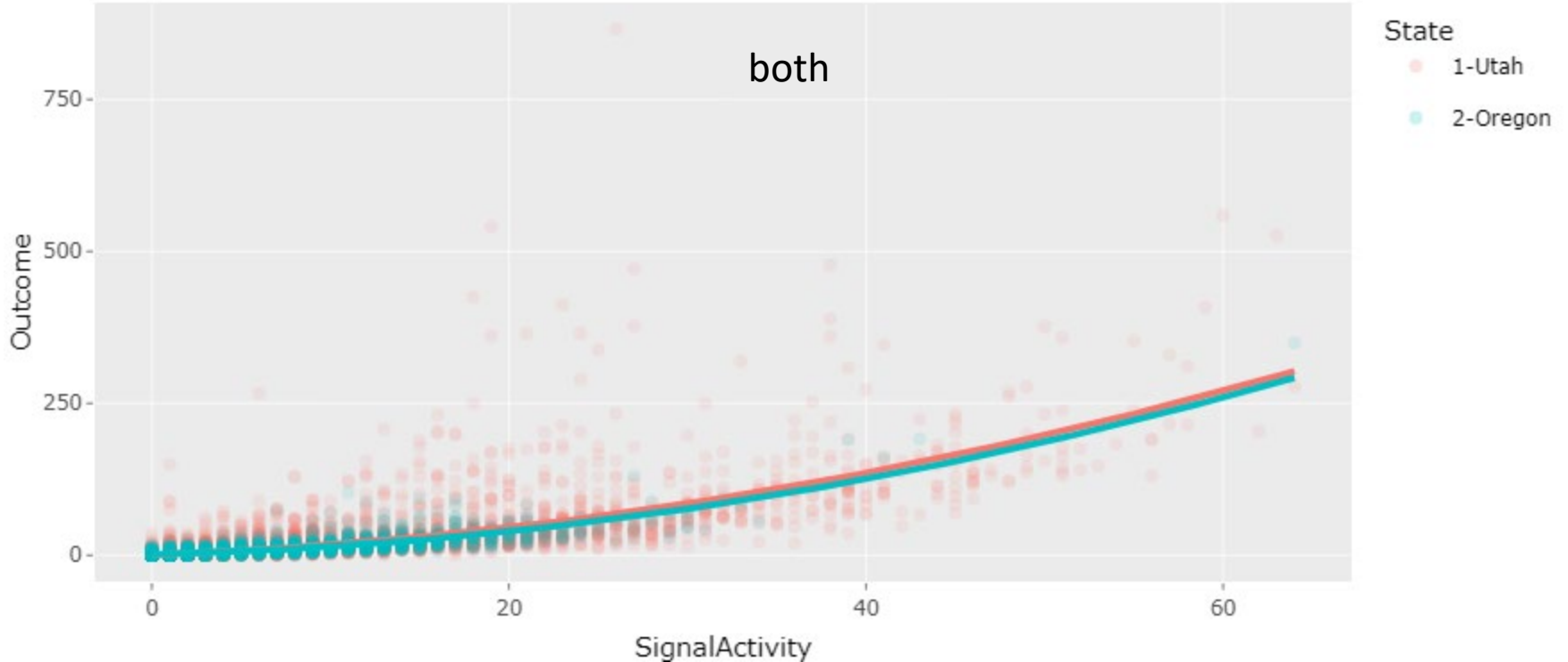
OR vs. UT

- How does Oregon data compare to Utah data?



OR vs. UT

- How does Oregon data compare to Utah data?



OR vs. UT

- Validation statistics for models estimated on Oregon and Utah data

Data	Model	COR	RMSE	MAE	SMAPE	MASE
Oregon	Oregon	0.8706	4.7477	2.3028	0.9456	0.7760
	Utah	0.8680	5.0195	2.3746	1.0321	0.8002
Utah	Utah	0.7685	14.8584	3.7847	1.1557	0.9500
	Oregon	0.7673	14.9890	3.7442	1.0916	0.9398

Conclusions

- Pedestrian pushbutton actuation data can be used for deriving pedestrian volumes at signalized intersections
- The recommended model demonstrated remarkable accuracy and generalizability, considering its simplicity
 - Dependent variable (TOTAL) is the total of all crosswalk users
 - Independent variable (A90C) is an imputed version of pedestrian detections using a 15 sec filter
 - Quadratic form, R^2 value of 0.76
 - The model's predictions were strongly correlated (0.86) with observations in 10% hold-out samples
 - Model's average absolute error (MAE) was ± 2.4 pedestrians per hour.
- Tests of transferability were promising and indicate that the model may be transferable to locations without similar data

Limitations

- Potential challenges associated with push-button data include data loss or corruption arising from equipment malfunction or communication issues
- The difference in signal operation limits the availability of push-button data
- The higher error observed when estimating high pedestrian volumes suggests that the model could benefit from more data collected at such locations to improve its accuracy
- It should also be noted that the seasonal differences in actuation and pedestrian volumes have not been explored yet

Recommendations

- Continuously archive pedestrian pushbutton actuation data as part of an ATSPM system to generate pedestrian volumes and other performance measures
 - If storage of raw data becomes an issue, larger time bins (e.g., 15-min) can be considered
- Perform regular validation of model outputs at a few locations to ensure model accuracy

Future work

- Investigate the source of errors during transferability and devise methods to mitigate them.
- Replicate study in other regions of the U.S. to understand the impact of regional variations.
- Investigate the impact of seasonal variation on pedestrian volume estimation.
- Explore the role of potential factors in pedestrian volume prediction, such as land uses, the built environment, roadway design, and weather.
- Investigate alternate methods of processing traffic signal controller log events.
- Explore the use of outlier detection techniques and missing data imputation methods to improve data accuracy.
- Investigate other statistical and machine learning models.
- Investigate the transferability of these pedestrian volume models to other jurisdictions.
- Explore how frequently should these models be re-estimated and under which circumstances.

Acknowledgements

- ODOT
- Josh Roll, Research Coordinator
- Members of the TAC
- Undergraduate students at PSU and USU

Bonus Slides

Data analysis principles

Accurate

- Predicts pedestrian volumes with low error and little-to-no bias (under/over-predict).

Generalizable

- Works just as well on out-of-sample data, from different places and/or times.

Simple

- Uses few inputs that are readily available and easily and consistently calculated.

Intuitive

- Variables and relationships have a straightforward and logical interpretation.

Data analysis principles

Accurate

- Measure quantitatively: Loglikelihood, LR Test, R^2 , Correlation, AIC, BIC, AICc, RMSE, MAE, SMAPE, and MASE.

Generalizable

- Use a 10-fold cross-validation. (90 % train vs. 10% test) \times 10 times.
- Compare to Utah.

Simple

- Use only one signal-related independent variable, and a few model segmentations.

Intuitive

- Must be able to easily explain results that make sense to everyone.

Data analysis approach

- Units of analysis
 - Time: 1 hour*
 - Space: ped phase # (1 crossing at 1 signal)*
- Dependent variable (DVs)
 - “Pedestrian” crosswalk crossing volume (count of users)
 - PED (PED only)
 - UPED* (PED+SKATE+WHEEL)
 - TOTAL (all crosswalk users)

Data analysis approach

- Independent variables (IVs)
 - Pedestrian signal activity measures
 - A45: # pedestrian call registered (event code 45)
 - A45A/B/C: # pedestrian actuations (imputed)
 - A45A: # times 90 after 0 or 22
 - A45B*: # times 90 after 0 or 21
 - A45C: # times 90 after 0
 - A90: # pedestrian detector on (event code 90)
 - A90A/B/C: # unique pedestrian detections (imputed)
 - A90A: # 90s at least 5 sec apart
 - A90B: # 90s at least 10 sec apart
 - A90C*: # 90s at least 15 sec apart

* Also used in Utah study.

Data analysis approach

- Functional forms (FFs)
 - Accommodate non-linearities in relationships
 - Candidate models: Linear, Piecewise, Quadratic, Cubic, Exponential, Power.
 - Included intercept in all the models
- Model segmentation/interactions (SIs)
 - RECALL*: Pedestrian recall
 - CYCLE090*: The average cycle length > 90 seconds
 - CYCLE120: The average cycle length > 120 seconds
 - PEAKAM: AM peak hours (7 and 8 AM)
 - PEAKPM: PM peak hours (16 and 17 PM)
 - PEAKAMP: Both the AM & PM peak hours

* Also used in Utah study.

Data analysis approach

- Model segmentation/interactions (SIs) (continued...)
 - OVERNIGHT: HOUR = 21, 22, 23, 0, 1, 2, 3, 4, 5
 - WEEKEND: DAY = Sat & Sun
 - ODOTREG: ODOT Region:
 - 1: Portland Metro, 2: Willamette Valley and North Coast, 3: Southwestern Oregon, 4: Central Oregon, 5: Eastern Oregon)
 - PLACETYPE3: ODOT place type, and combined into 3 categories:
 - Low or non-MPO: Rural, Rural Near Major Center, MPO Low Density, Isolated City, City near Major Center;
 - Medium MPO (MPO Residential, MPO Employment); and
 - High MPO (MPO Mixed Use, MPO TOD)
 - TRANSIT4C: any transit stops (within 400 m)
 - TRANSIT8C: any transit stops (within 800 m)
 - EDTOTAL4C: any educational institutions (within 400 m)
 - EDTOTAL8C: any educational institutions (within 800 m)

Data analysis approach

- Model validation
 - K-fold cross-validation, splitting the dataset into 10 sections
 - 90% training dataset, 10% testing dataset
 - Random selection of signals
 - Each fold consists of 6-7 signals data
 - Goodness-of-fit measures:
 - Log-likelihood, likelihood ratio test, R^2 , correlation, AIC, BIC, and AICc.
 - Accuracy measures:
 - RMSE, MAE, SMAPE, and MASE.