



Oregon Department of Transportation



ODOT Research

Methods for Monitoring Nonmotorized Transportation - A Proof of Concept in Bend, OR

6/17/2020



Background

Josh's Background

- Travel, land use, air quality, and GHG modeling
- Traffic count program development
- Crash safety analysis
- Public health analysis

ODOT Role

- Active and Sustainable Transportation Research Coordinator
- Coordinate and conduct research



Oregon Department of Transportation



Agenda



Agenda

Background/Objectives

Why Count
Nonmotorists?

Count Program

Data Fusion Modeling

Next Steps

Discussion & Questions

Research Objectives

Initial Objectives

- Assist Bend MPO in setting up multimodal traffic data collection system
 - Measure project success
 - Plan for the future
 - Prioritize maintenance activities and operations
 - Improve safety analysis
- Measure crash risk for all modes

High Level Objectives

- Develop data collection system with ability to scale easily to other urban areas
- Make it simple and automated as possible
- Provide usable data for high end uses (planning modeling, KPM, health analysis)



Why Count Nonmotorized Traffic?

Invisible Traffic

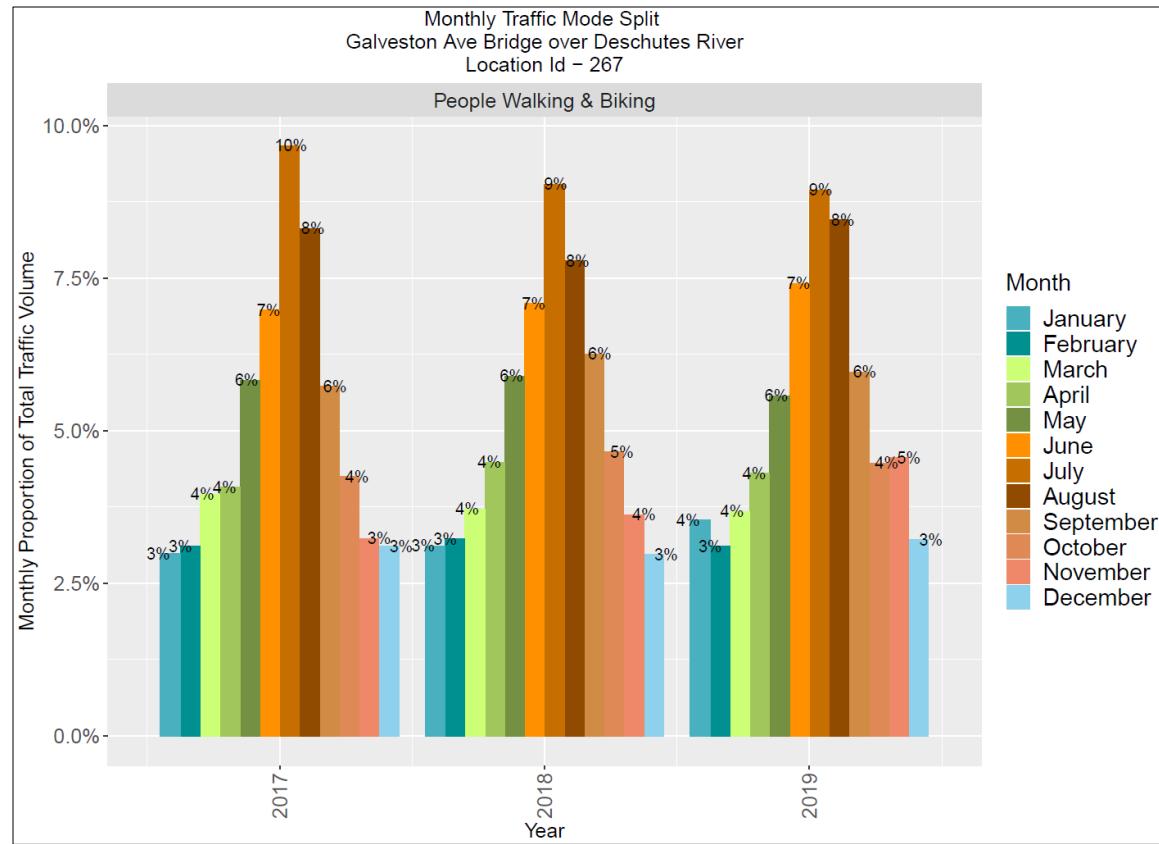
- What's not counted doesn't count
- Short term counts not the whole picture

Highlighting Invisible Traffic

- 405K Vehicle Traffic (July)
- 41K Bike & Pedestrian Traffic

Modal Comparisons

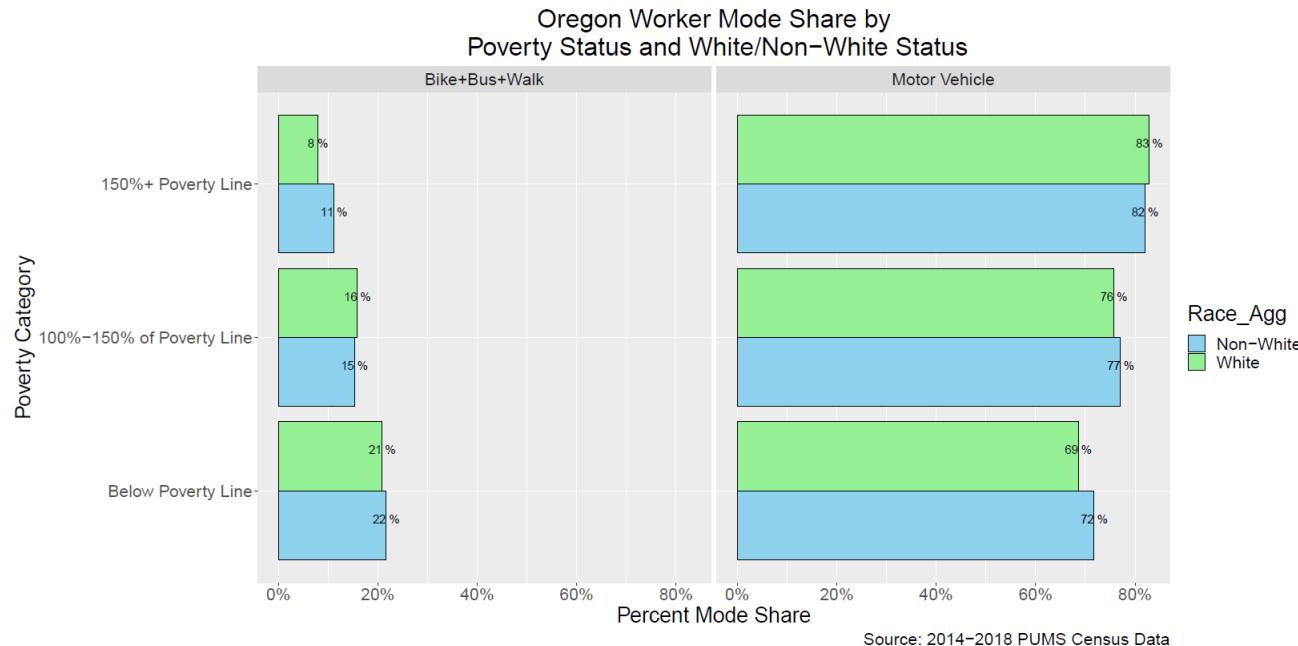
- Segment mode share not a static property



Why Count Nonmotorized Traffic?

Social Equity

- Social justice implications for now accounting for nonmotorized traffic activity



Count Program Overview

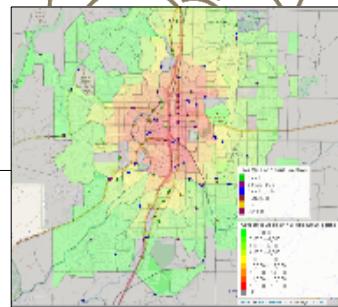
Field Staff Data Entry



Eco Counter
Counts Database

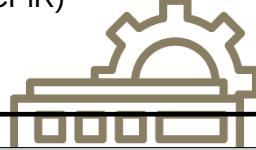


Cloud Based Data Storage

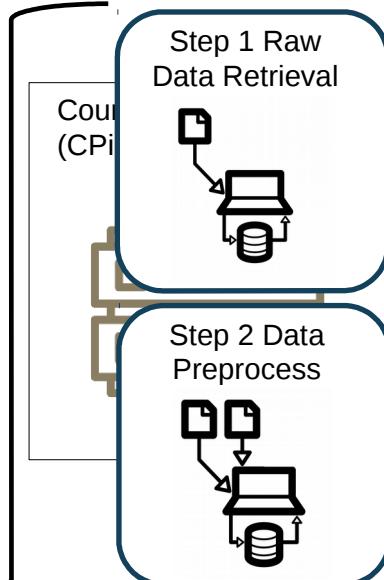


Count Station
Spatial Data

Counts Processor in R
(CPiR)


1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	10010	10011	10012	10013	10014	10015	10016	10017	10018	10019	10020	10021	10022	10023	10024	10025	10026	10027	10028	10029	10030	10031	10032	10033	10034	10035	10036	10037	10038	10039	10040	10041	10042	10043	10044	10045	10046	10047	10048	10049	10050	10051	10052	10053	10054	10055	10056	10057	10058	10059	10060	10061	10062	10063	10064	10065	10066	10067	10068	10069	10070	10071	10072	10073	10074	10075	10076	10077	10078	10079	10080	10081	10082	10083	10084	10085	10086	10087	10088	10089	10090	10091	10092	10093	10094	10095	10096	10097	10098	10099	100100	100101	100102	100103	100104	100105	100106	100107	100108	100109	100110	100111	100112	100113	100114	100115	100116	100117	100118	100119	100120	100121	100122	100123	100124	100125	100126	100127	100128	100129	100130	100131	100132	100133	100134	100135	100136	100137	100138	100139	100140	100141	100142	100143	100144	100145	100146	100147	100148	100149	100150	100151	100152	100153	100154	100155	100156	100157	100158	100159	100160	100161	100162	100163	100164	100165	100166	100167	100168	100169	100170	100171	100172	100173	100174	100175	100176	100177	100178	100179	100180	100181	100182	100183	100184	100185	100186	100187	100188	100189	100190	100191	100192	100193	100194	100195	100196	100197	100198	100199	100200	100201	100202	100203	100204	100205	100206	100207	100208	100209	100210	100211	100212	100213	100214	100215	100216	100217	100218	100219	100220	100221	100222	100223	100224	100225	100226	100227	100228	100229	100230	100231	100232	100233	100234	100235	100236	100237	100238	100239	100240	100241	100242	100243	100244	100245	100246	100247	100248	100249	100250	100251	100252	100253	100254	100255	100256	100257	100258	100259	100260	100261	100262	100263	100264	100265	100266	100267	100268	100269	100270	100271	100272	100273	100274	100275	100276	100277	100278	100279	100280	100281	100282	100283	100284	100285	100286	100287	100288	100289	100290	100291	100292	100293	100294	100295	100296	100297	100298	100299	100300	100301	100302	100303	100304	100305	100306	100307	100308	100309	100310	100311	100312	100313	100314	100315	100316	100317	100318	100319	100320	100321	100322	100323	100324	100325	100326	100327	100328	100329	100330	100331	100332	100333	100334	100335	100336	100337	100338	100339	100340	100341	100342	100343	100344	100345	100346	100347	100348	100349	100350	100351	100352	100

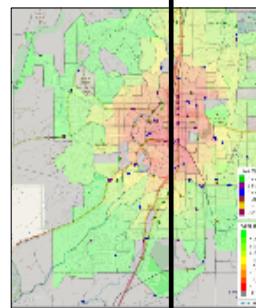
Count Program Overview



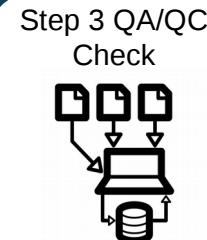
- Pull data
- Assign temporal and spatial info
- Assign detailed spatial data
- Process portable sites
- Split user counts



Eco Counter
Counts Database



Count Station
Spatial Data



Assign error flags

Google Street
Repository

Daily count imputation
Annual estimates applying Doy factors



Deployment
Picture



We Have Counts Data....Now What?

Goal

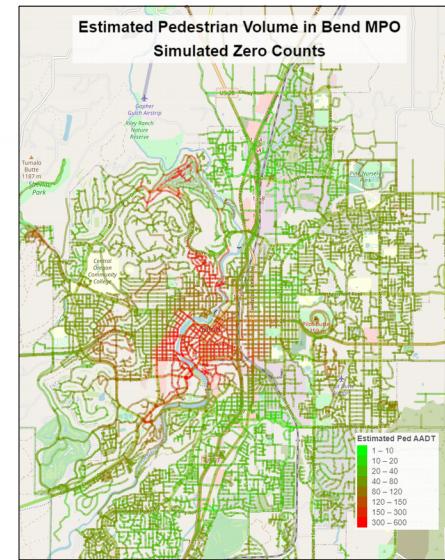
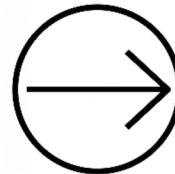
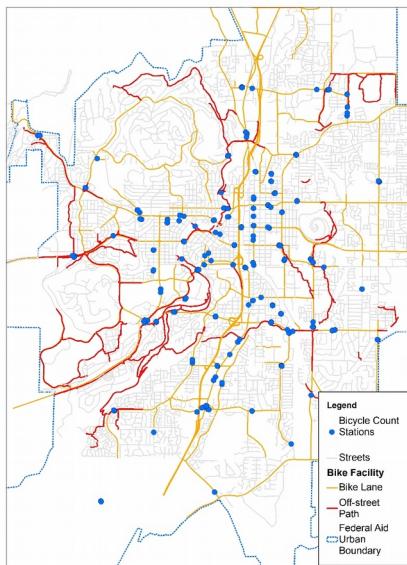
- Estimate activity across the network

Issue: Limited Spatial Resolution

- 56 - 94 sites

Solution: Model traffic

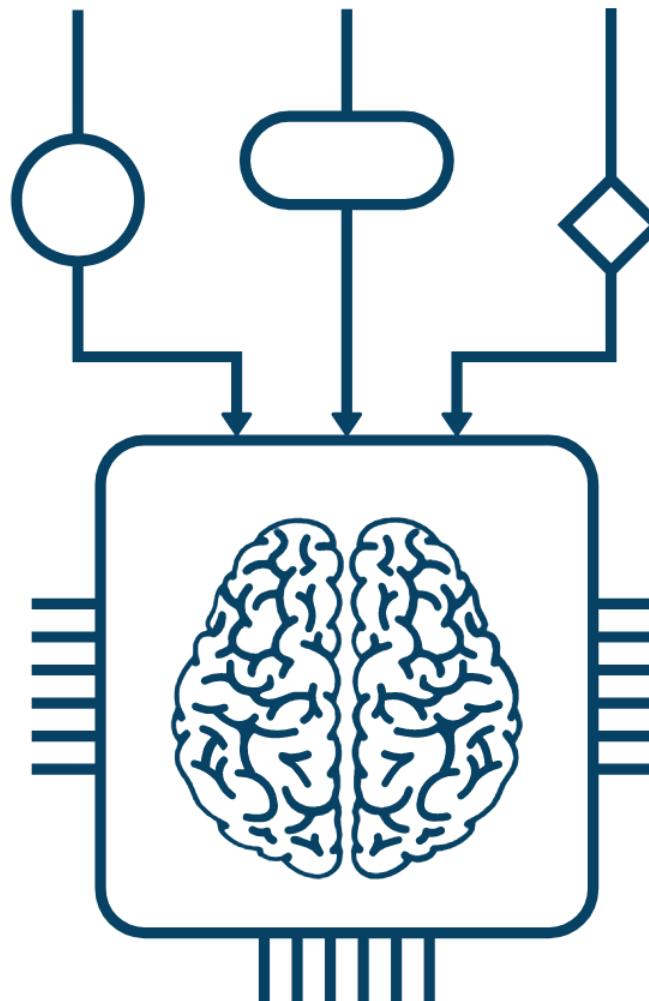
- Use relationships between land use, accessibility and network features and counts
- Parametric vs. machine learning approaches



Data Fusion with Machine Learning

What is Machine Learning?

- Algorithms that find and apply patterns in data (MIT Technology Review)
- Many different types for different purposes
- Classification vs. Regression
- Supervised vs. Unsupervised



Data Fusion with Machine Learning

Typical Uses

- Marketing, genetic research, physics, social media, and transportation!

Selected Methods

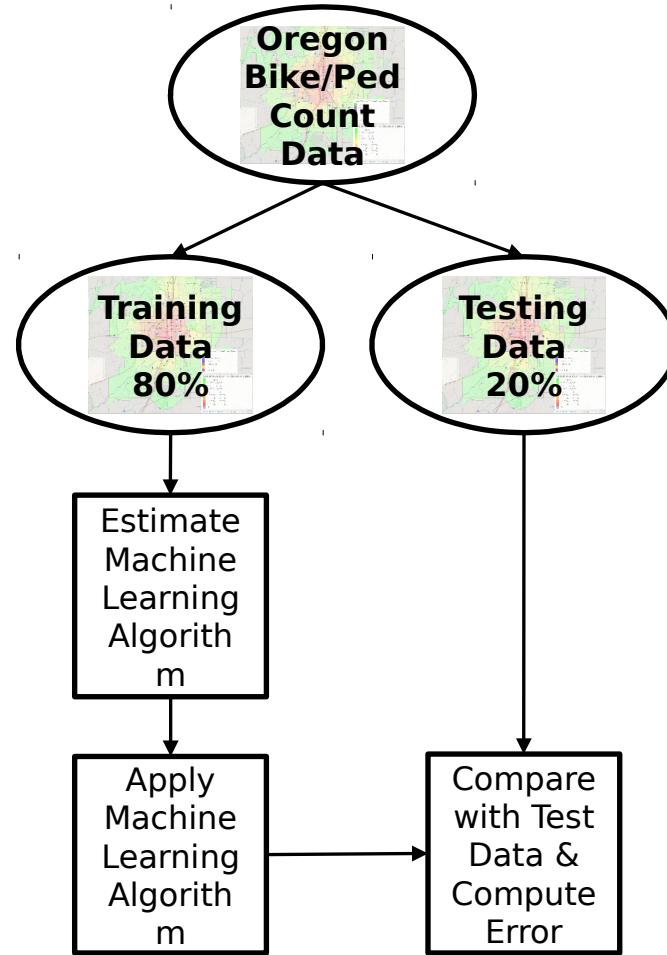
- Negative Binomial Regression
- Decision Tree
- Random Forest



Cross-Validation

Cross Validation

- Divide data into training and testing sets
- Training data for estimating model
- Testing data for determining accuracy of model
- Performed many times to ensure model stability



Network Modeling - Data Fusion

What is a model?

- Representation of a thing or phenomenon useful for understanding and decision making
- Performance of a model depends on uses and decisions being made
- “All models are wrong, some are useful”
- Data driven models allow us to put our assumptions on the table

Travel models poor tools for nonmotorized transportation

- Travel surveys collect limited information on nonmotorized
- Assignment procedures make oversimplified assumptions
- No bike/ped counts to calibrate to anyway
- TDMs been a little tyrannical



Network Modeling - Data Fusion

Objective

- Activity estimates for entire network

Uses

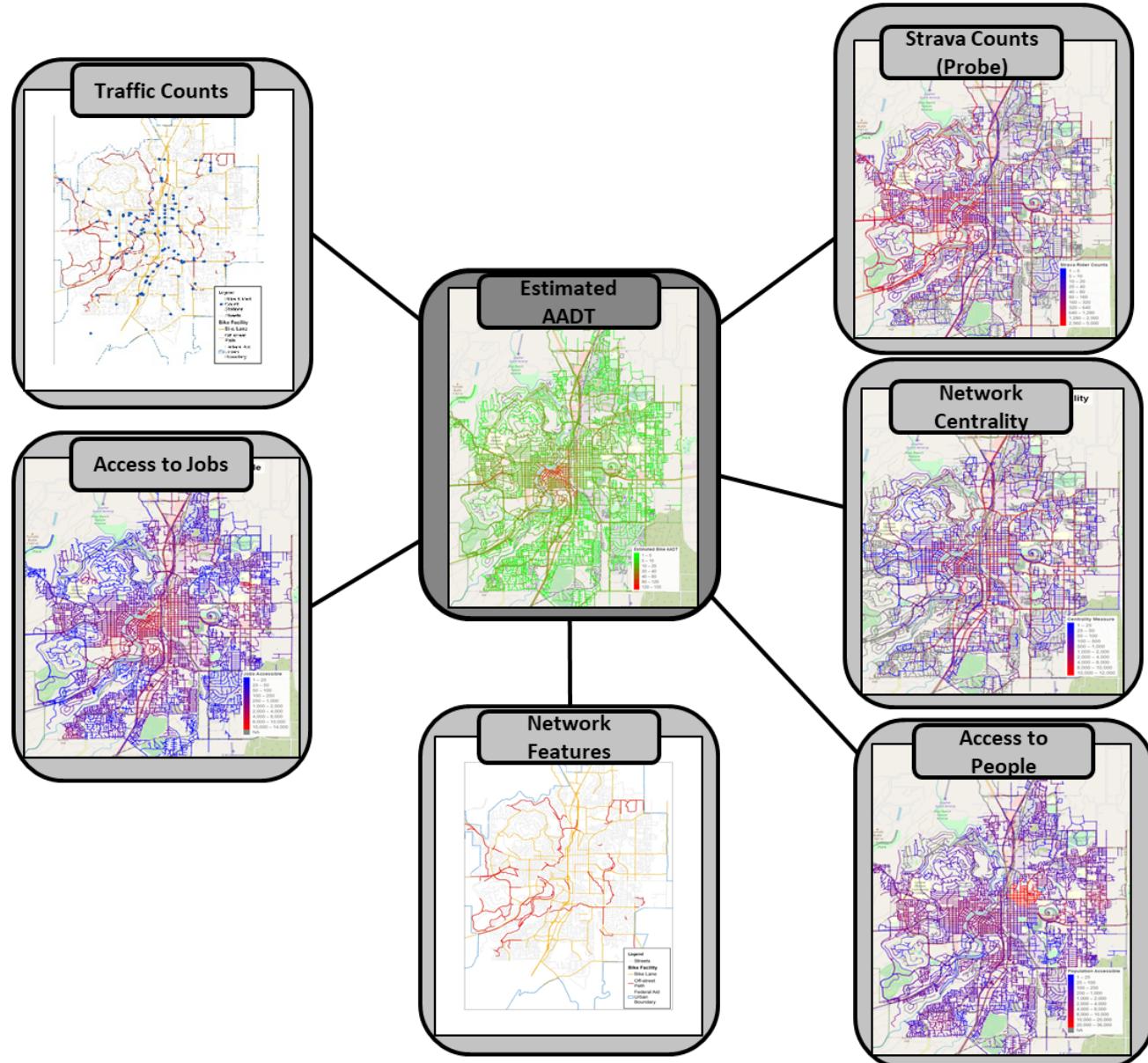
- Planning, monitoring, crash analysis

Methods

- Merges data from multiple features and apply machine learning or statistical model

Output

- Quantifying total network activity
- Crash analysis input
- Health analysis input



Network Modeling - Data Fusion

User Types

- Vehicle
- Bicycle
- Pedestrian

Data

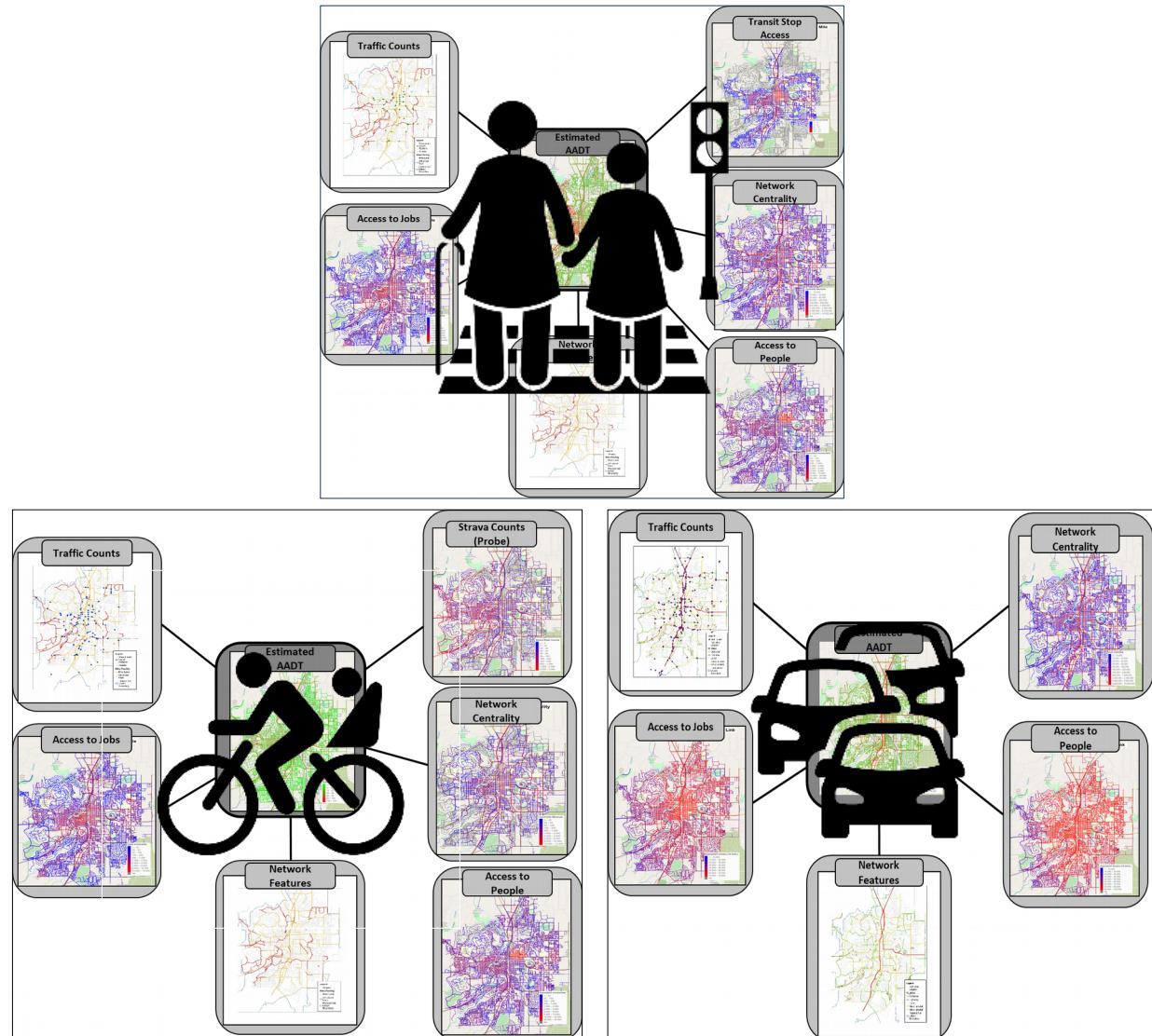
- Network characteristics
- Accessibility
- Centrality
- Probes

Methods

- Random forest and XgBoost

Output

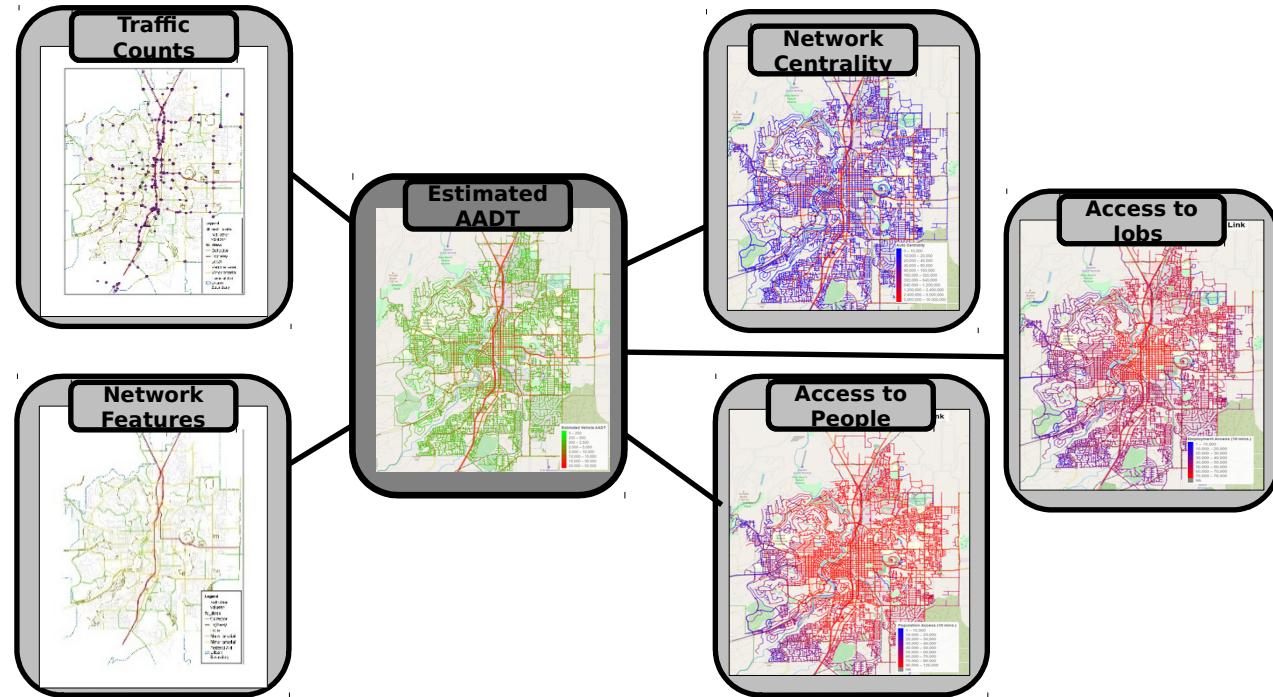
- Quantifying total network activity
- Crash analysis input
- Health analysis input





Vehicle AADT Data Fusion Scheme

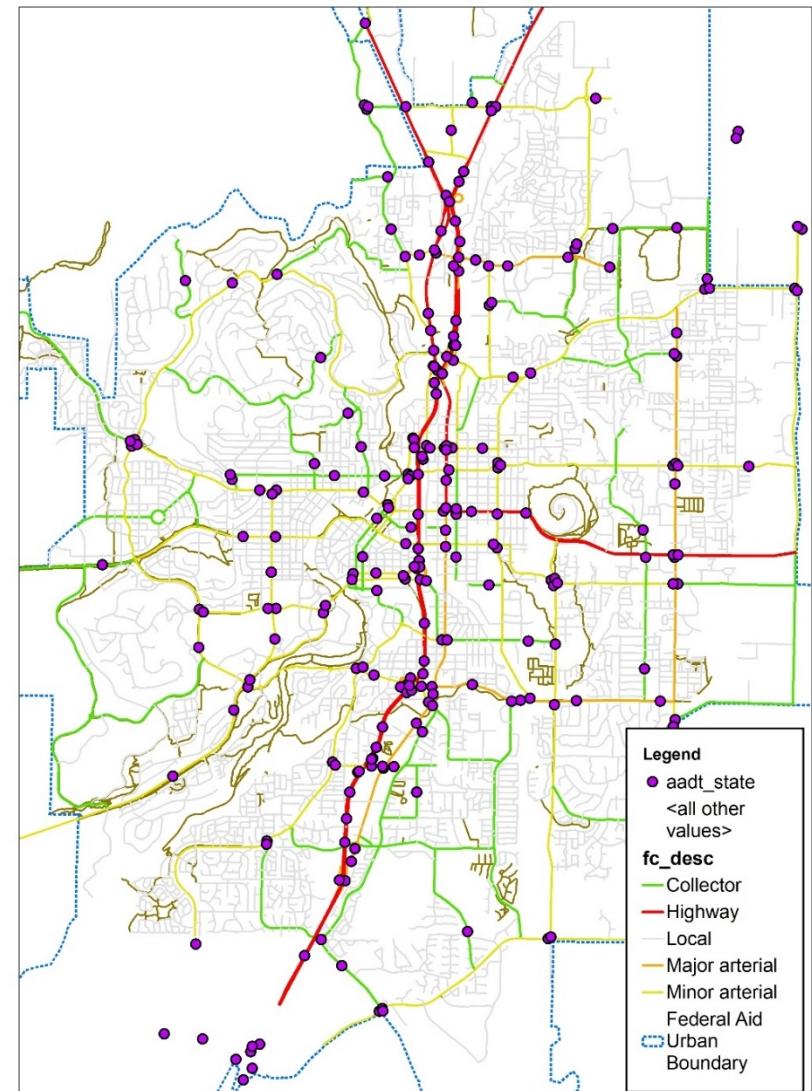
- Vehicle Model Objectives
 - Validate data fusion approach
 - Provide network wide estimates of vehicle traffic
- Data and Models Used
 - Up to 433 data features in some specs
 - XgbBoost & Random Forest
 - Census, TAZ, properly attributed routable network
- Validation
 - Internal 10-fold cross validations (random partitions)
 - External 10-fold (stratified partition)
 - Leave-one-out validation
 - Comparison with Federally reported data (HPMS)





Vehicle AADT Model Data

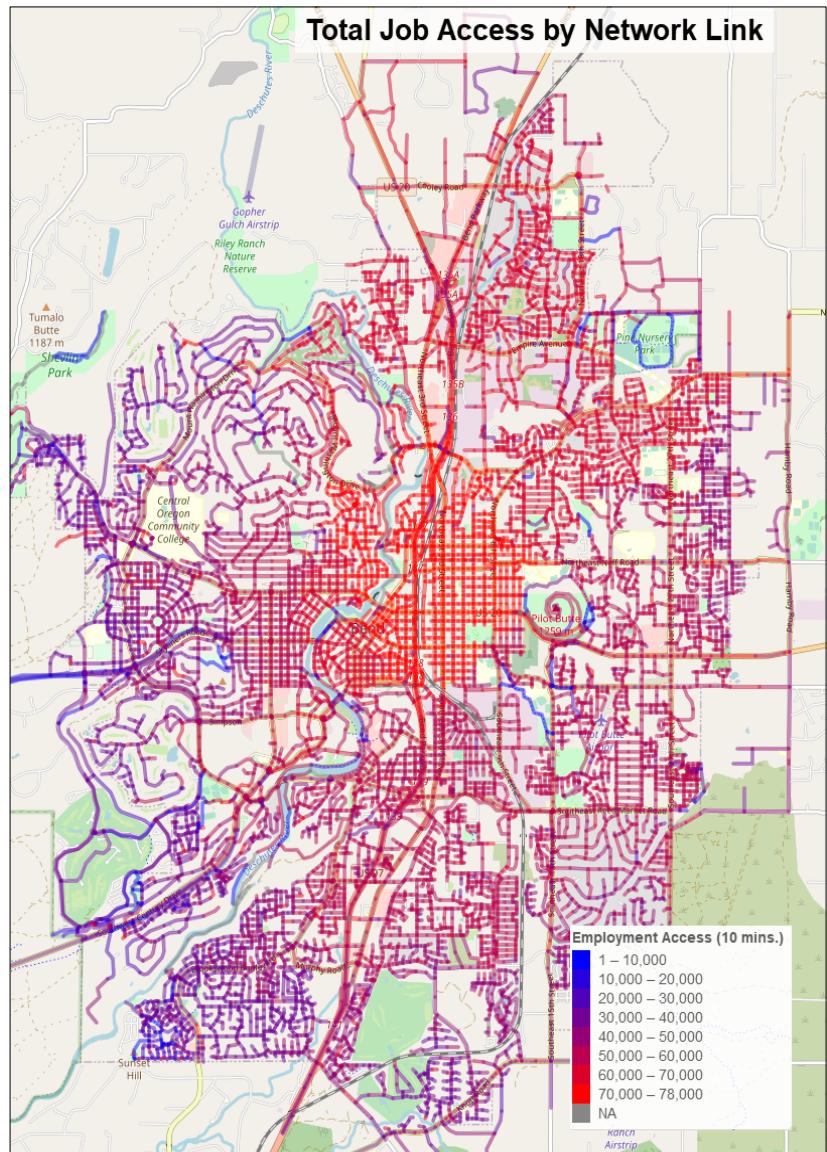
- Vehicle Model Data
 - Traffic Counts
 - 2018 & 2019
 - N = 255
 - Network Features
 - Functional classification
 - Posted speed limit





Vehicle AADT Model Data

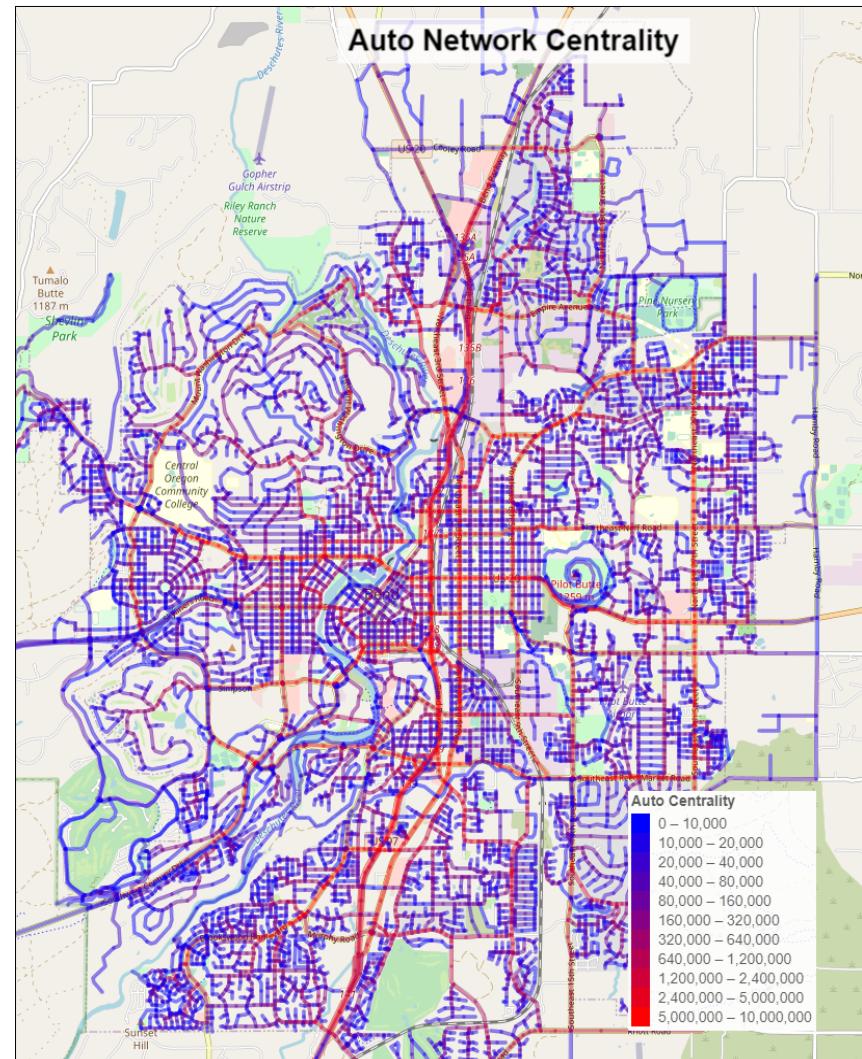
- Vehicle Model Data
 - Traffic Counts
 - 2018 & 2019
 - N = 255
 - Network Features
 - Functional classification
 - Posted speed limit
 - Accessibility (drive time)
 - Jobs
 - People





Vehicle AADT Model Data

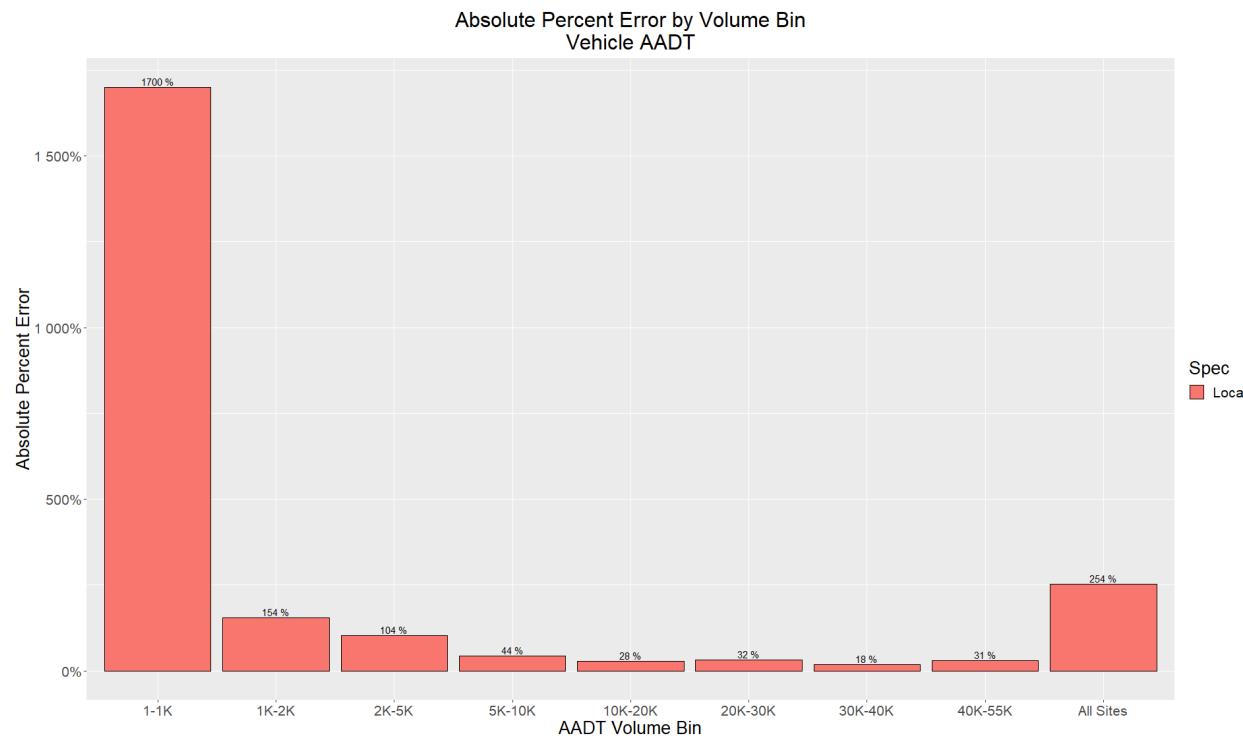
- Vehicle Model Data
 - Traffic Counts
 - 2018 & 2019
 - N = 255
 - Network Features
 - Functional classification
 - Posted speed limit
 - Accessibility
 - Jobs
 - People
 - Centrality
 - Measures link importance





Vehicle AADT Model Validation

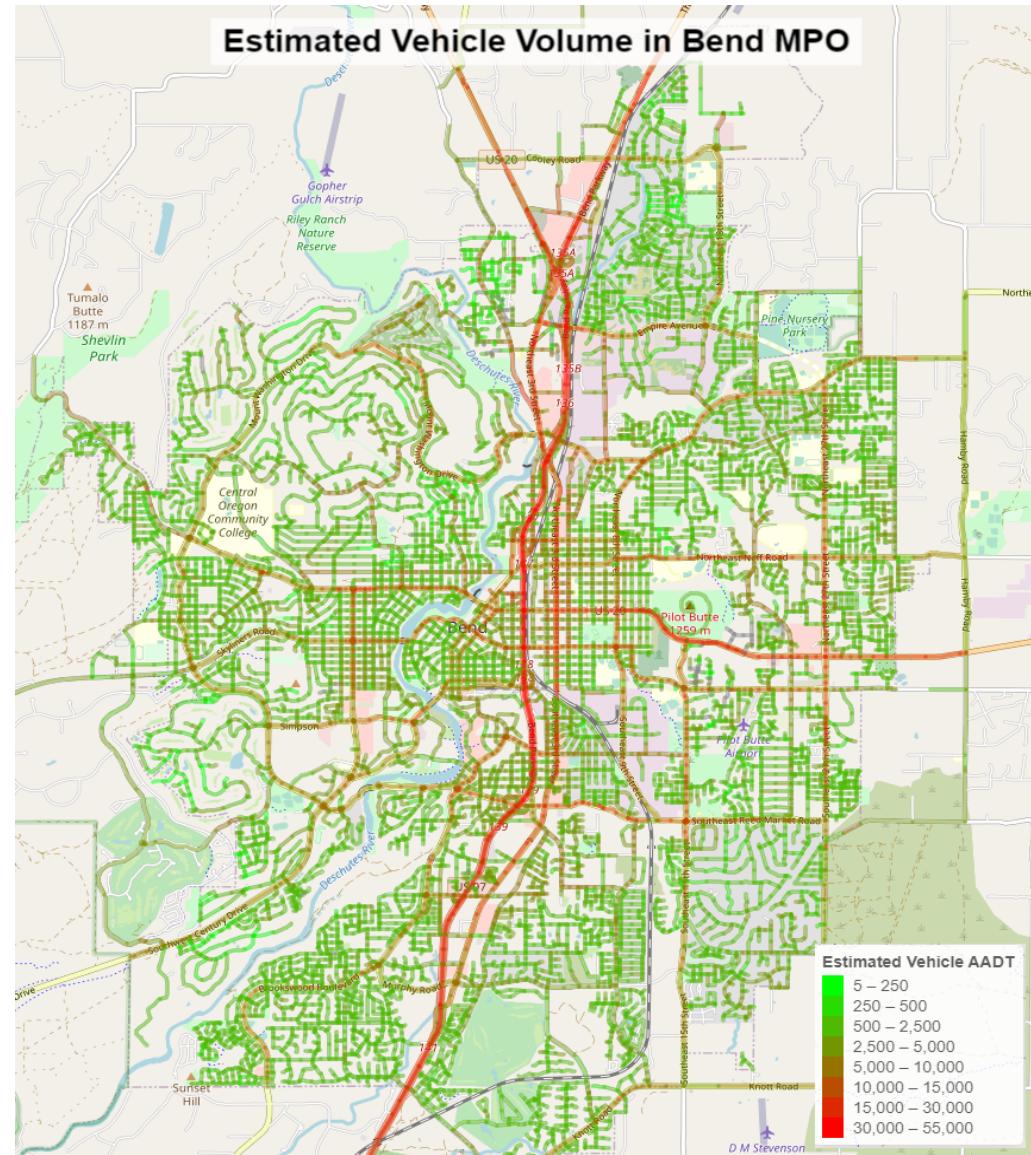
- 10-fold Cross-validation
 - Multiple specifications tried
 - local and federal fc
 - Prediction error varies by volume bin
 - Overall 254% error
 - 25% median error for volume bins 5K and greater





Vehicle AADT Model Results

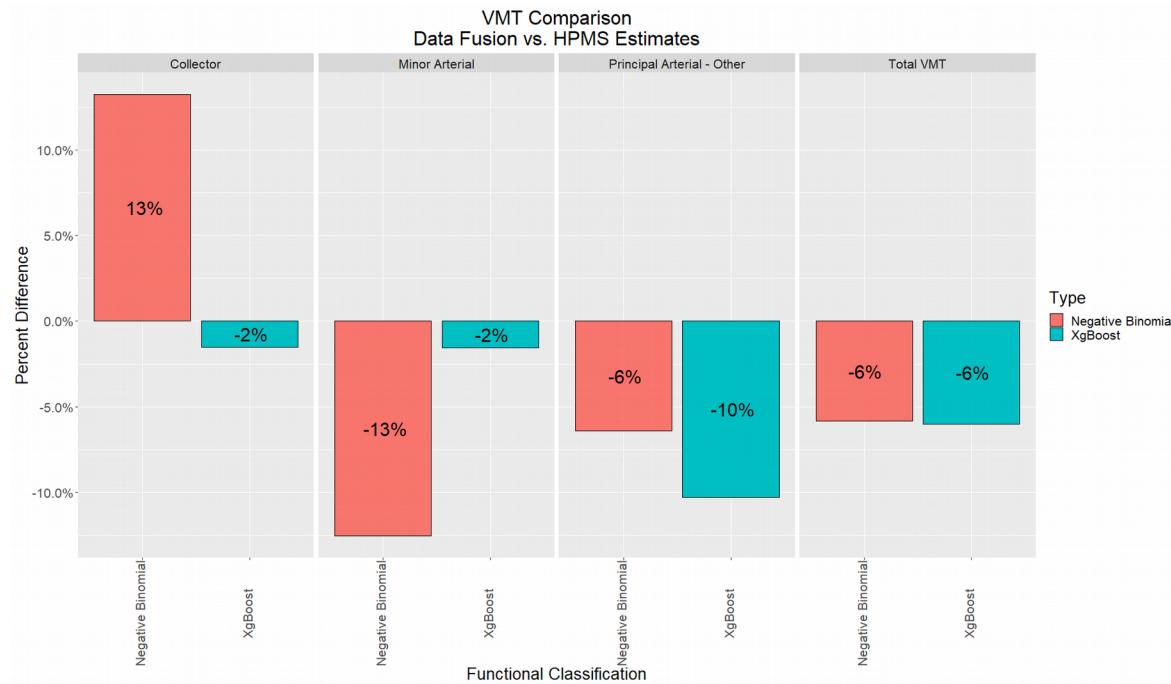
- Network wide estimates
 - High volume roads appear reasonable
 - Low volume local streets appear overestimated





Vehicle AADT Model Results Comparison

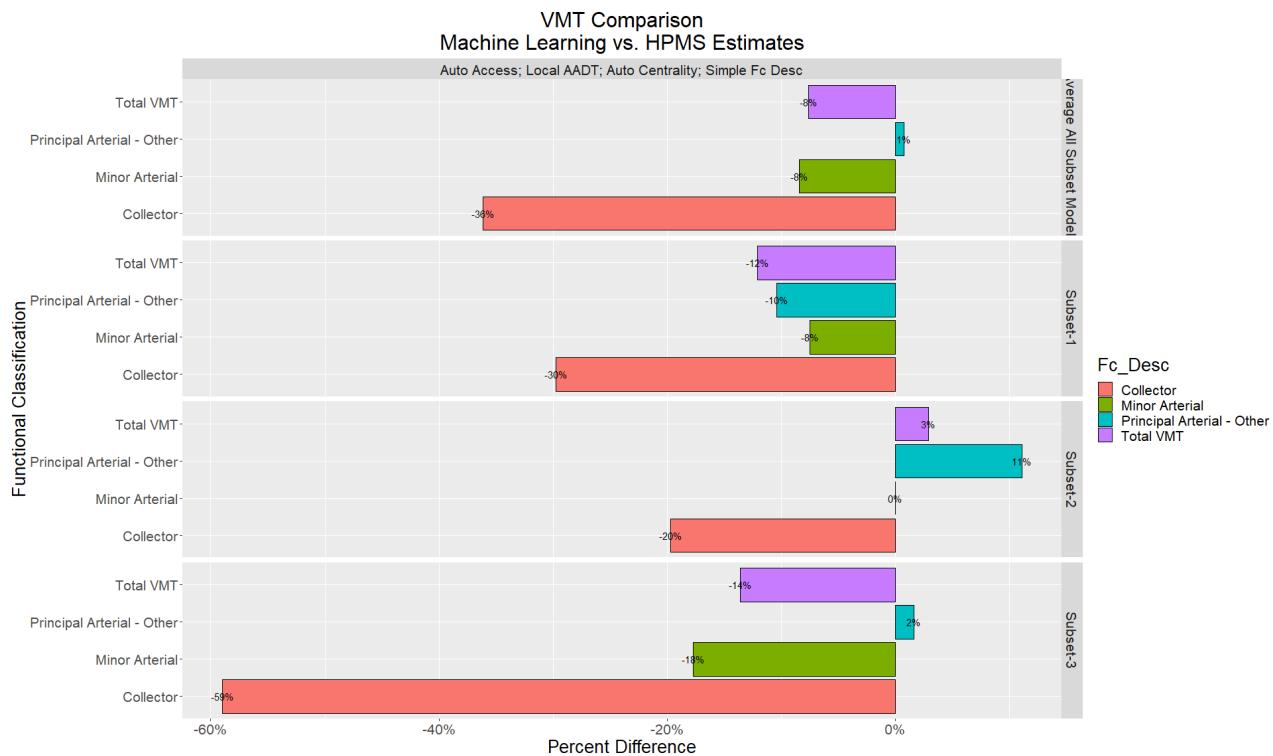
- Comparison with HPMS
 - Overall VMT estimate within 6% (model over estimates)
 - Model approaches provide reasonable system level estimates
 - Principal arterial highest error at 10% for ML
 - Collector & Min. Art. Highest error for Neg. Bin





Vehicle AADT Model Results Comparison

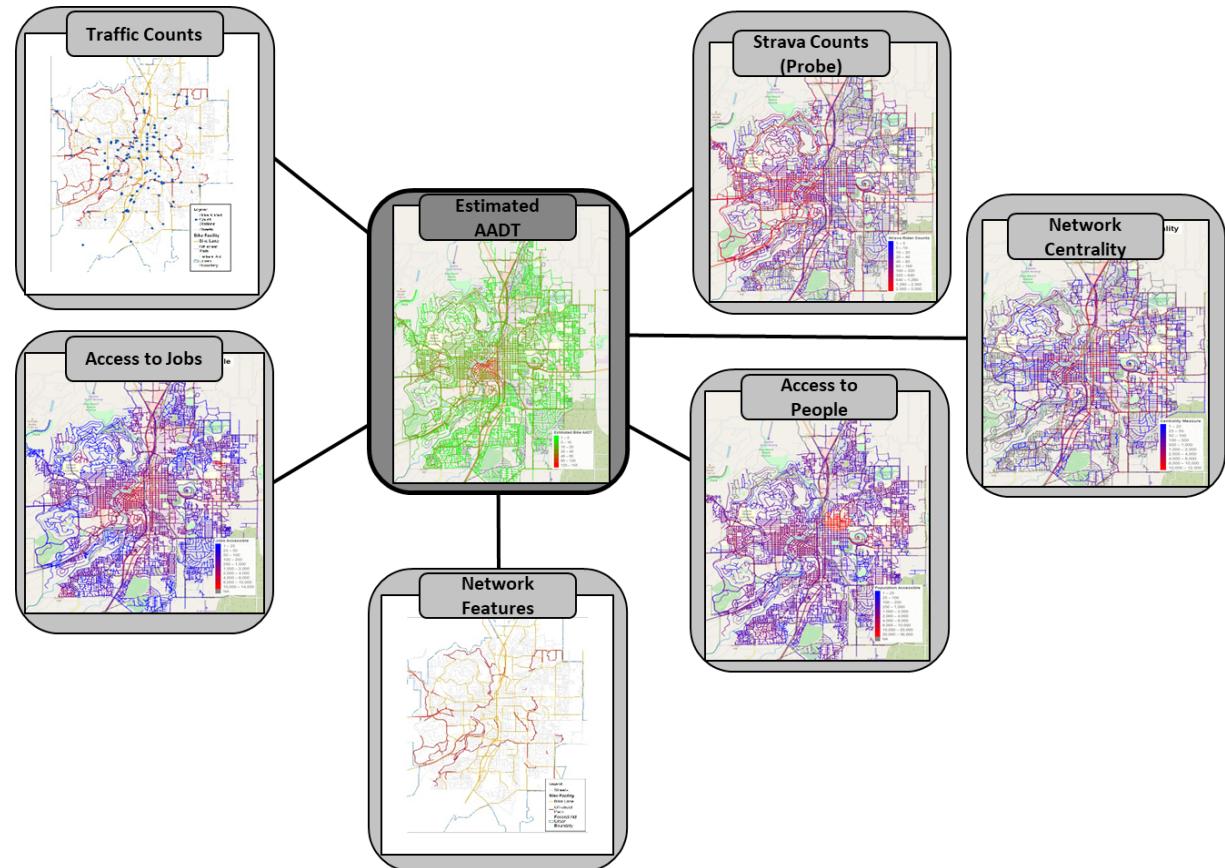
- Comparison with HPMS
 - Subset models are randomly partitioned into 3 datasets
 - Models within 3% to 14% compared to HPMS
 - Collectors perform poorly, likely due to small number of observations in training data
- Vehicle Model Conclusions
 - Approach performs well for aggregate and slightly disaggregate
 - Subset models improve confidence in
 - Disaggregate level useful in planning applications (& crash analysis?)
 - Results for each year available
 - Probe data will vastly improve approach (coming?)





Bicycle AADT Data Fusion Scheme

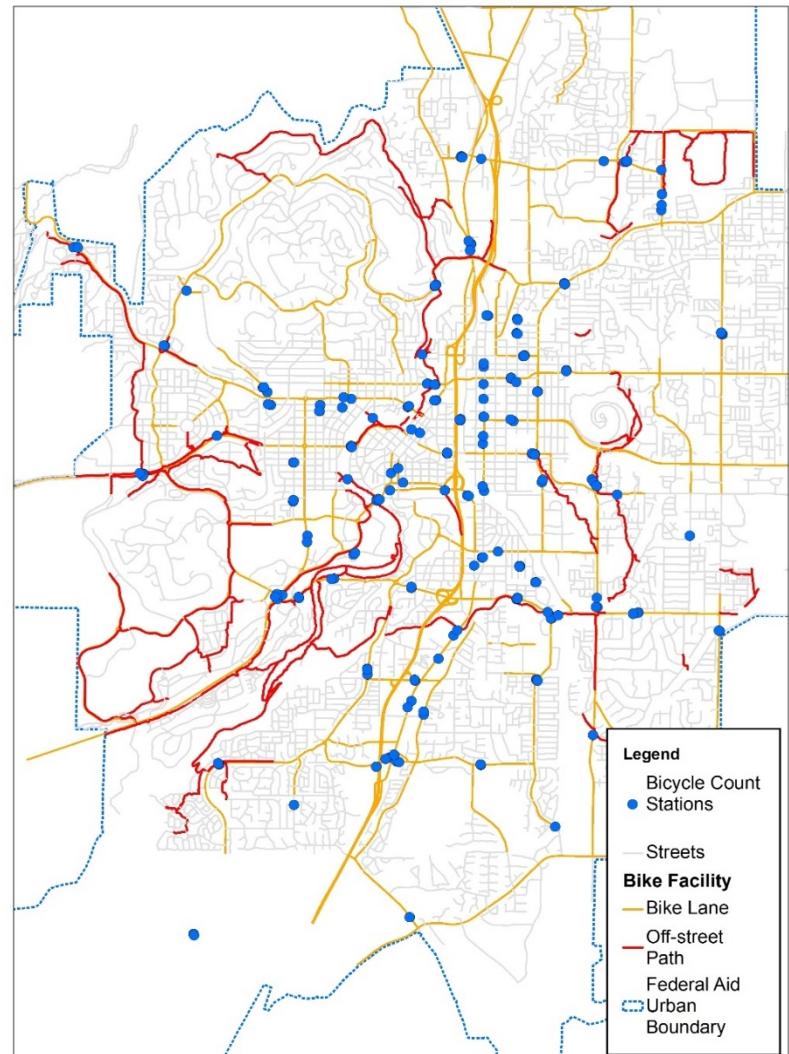
- Bicycle Model Objectives
 - Provide network wide estimates of bicycle traffic
- Data and Models Used
 - Up to 516 data features in some specs
 - XgbBoost & Random Forest
 - Census, TAZ, properly attributed routable network, and probe data
- Validation
 - Internal 10-fold cross validations (random partitions)
 - External 10-fold (stratified partition)
 - Leave-one-out validation





Bicycle AADT Model Data

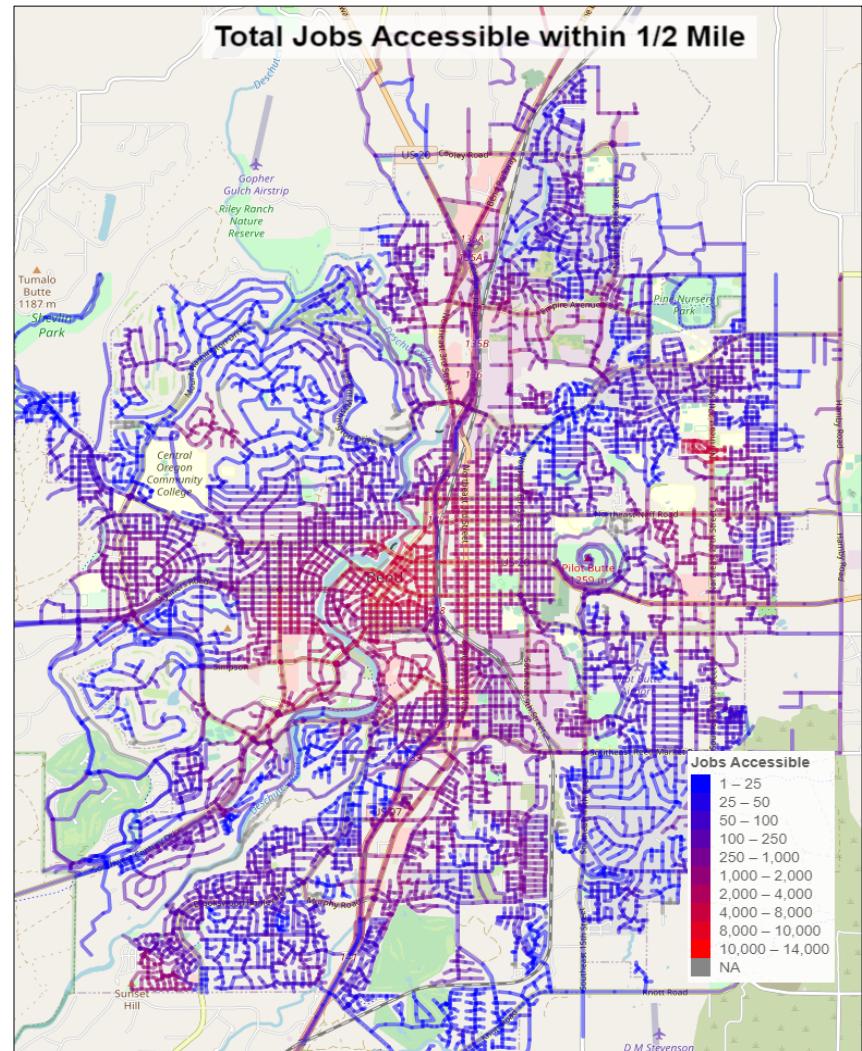
- Bicycle Model Data
 - Traffic Counts
 - 2017, 2018 & 2019
 - N = 94
 - Network Features
 - Functional classification
 - Posted speed limit
 - Bicycle facility type





Bicycle AADT Model Data

- Bicycle Model Data
 - Traffic Counts
 - 2017, 2018 & 2019
 - N = 94
 - Network Features
 - Functional classification
 - Posted speed limit
 - Bicycle facility type
 - Accessibility (distance)
 - Jobs
 - People

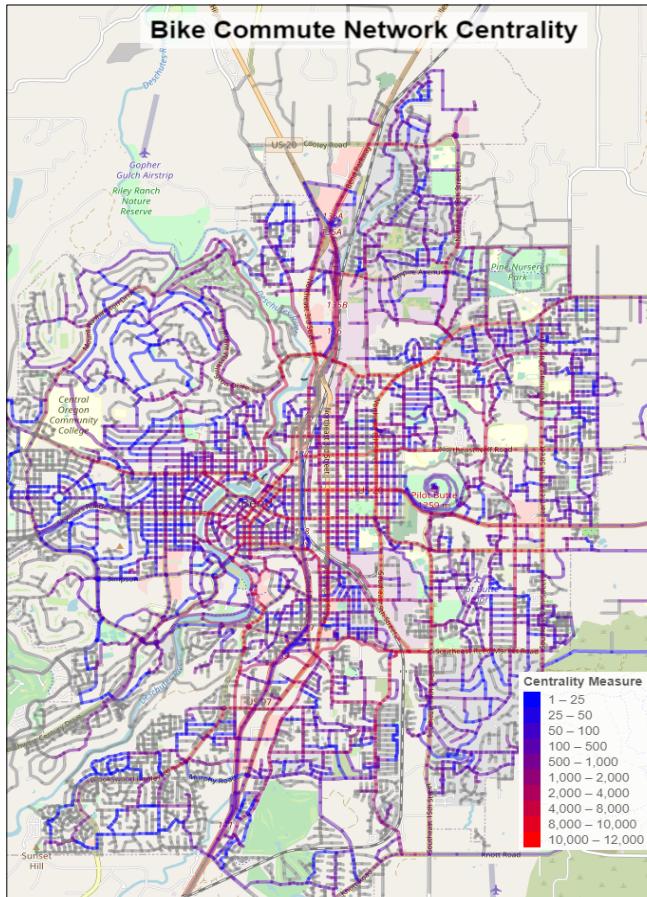




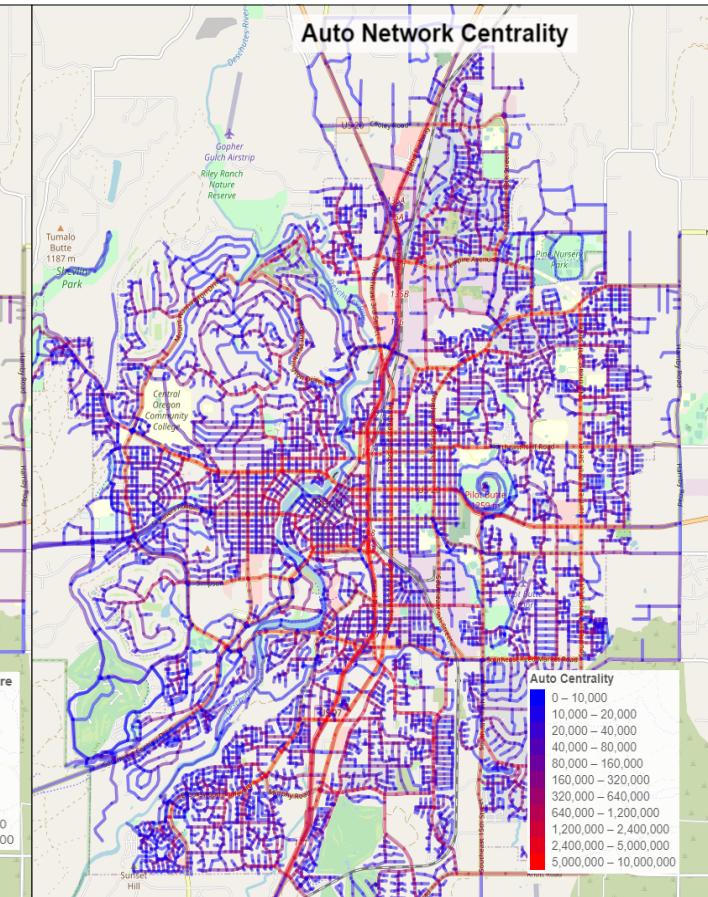
Bicycle AADT Model Data

- Bicycle Model Data
 - Traffic Counts
 - 2017, 2018 & 2019
 - N = 94
 - Network Features
 - Functional classification
 - Posted speed limit
 - Bicycle facility type
 - Centrality
 - Commute
 - Recreational
 - Accessibility (distance)
 - Jobs
 - People

Bicycle Centrality



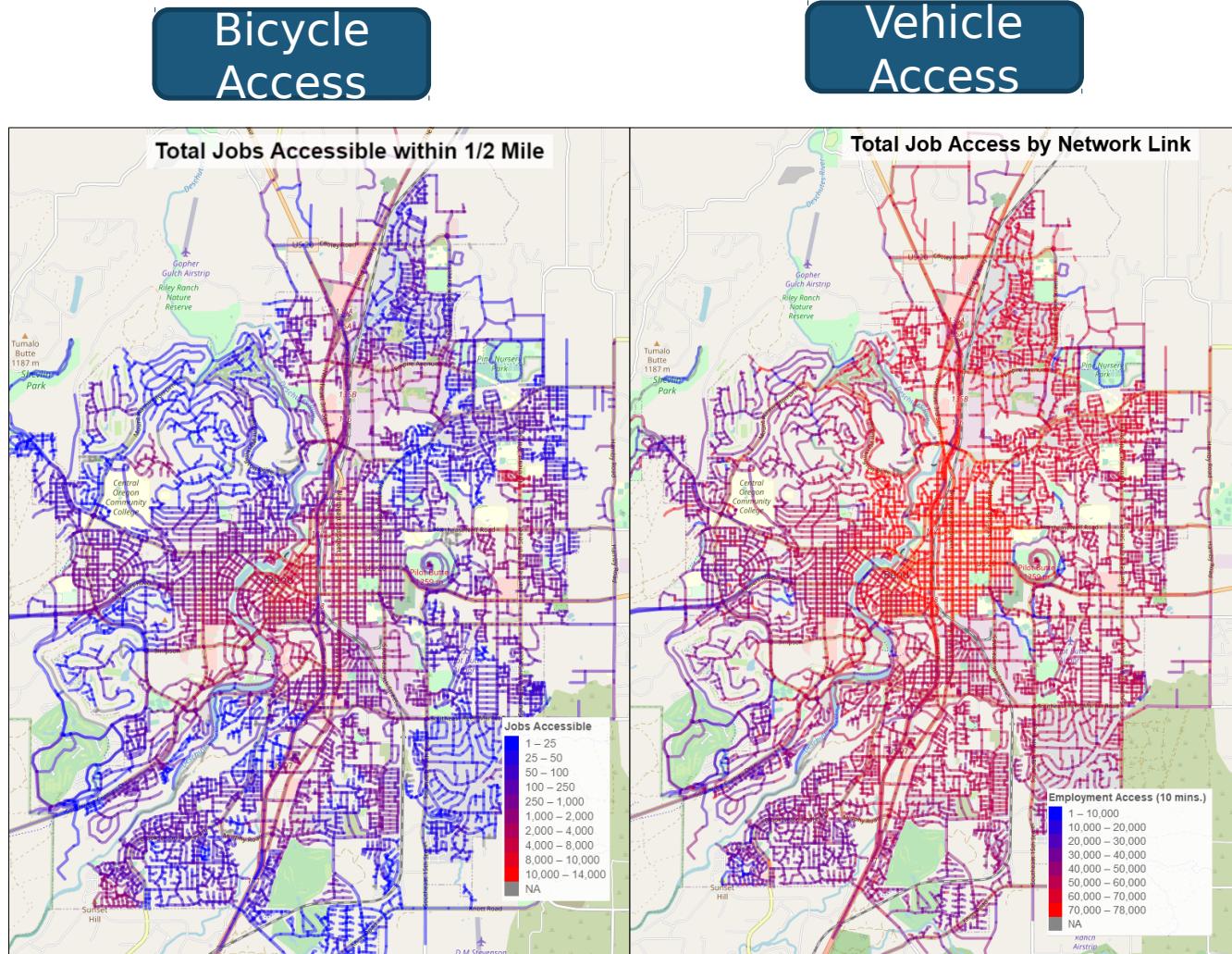
Vehicle Centrality





Bicycle AADT Model Data

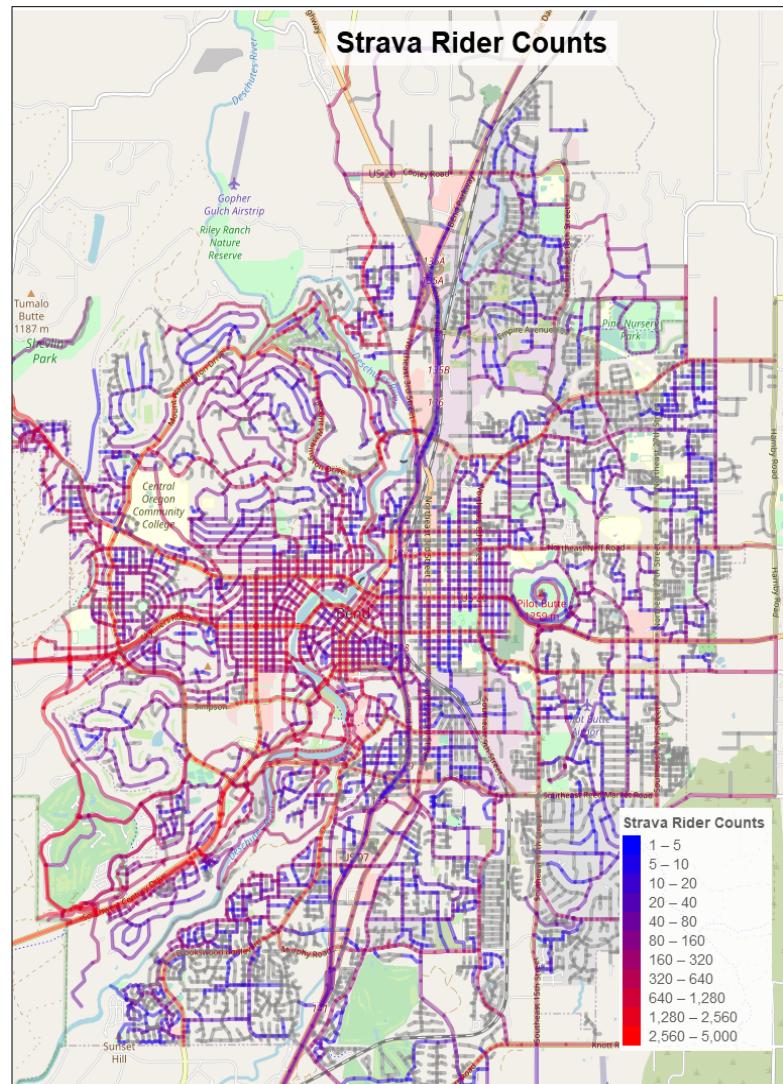
- Bicycle Model Data
 - Traffic Counts
 - 2017, 2018 & 2019
 - N = 94
 - Network Features
 - Functional classification
 - Posted speed limit
 - Bicycle facility type
 - Accessibility (distance)
 - Jobs
 - People





Bicycle AADT Model Data

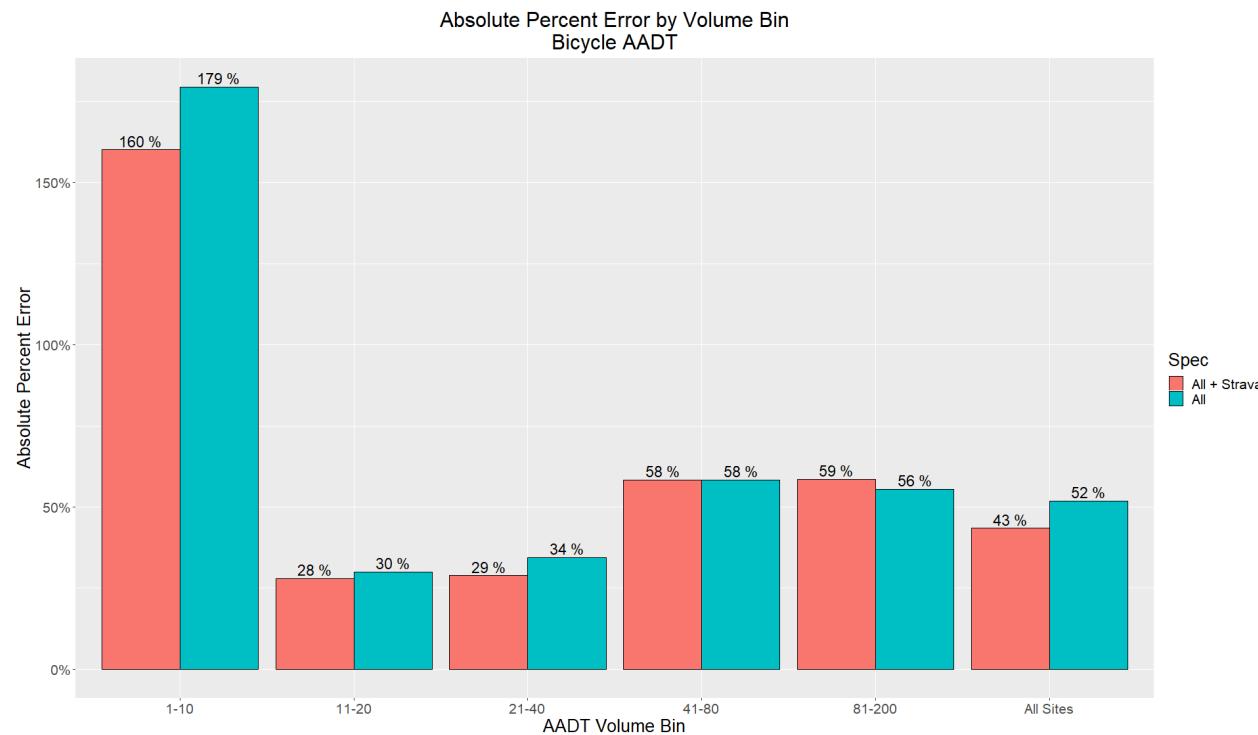
- Bicycle Model Data
 - Traffic Counts
 - 2017, 2018 & 2019
 - N = 94
 - Network Features
 - Functional classification
 - Posted speed limit
 - Bicycle facility type
 - Centrality
 - Commute
 - Recreational
 - Accessibility (distance)
 - Jobs
 - People
 - Probe Data
 - Strava
 - 2017-2019 data





Bicycle AADT Model Validation

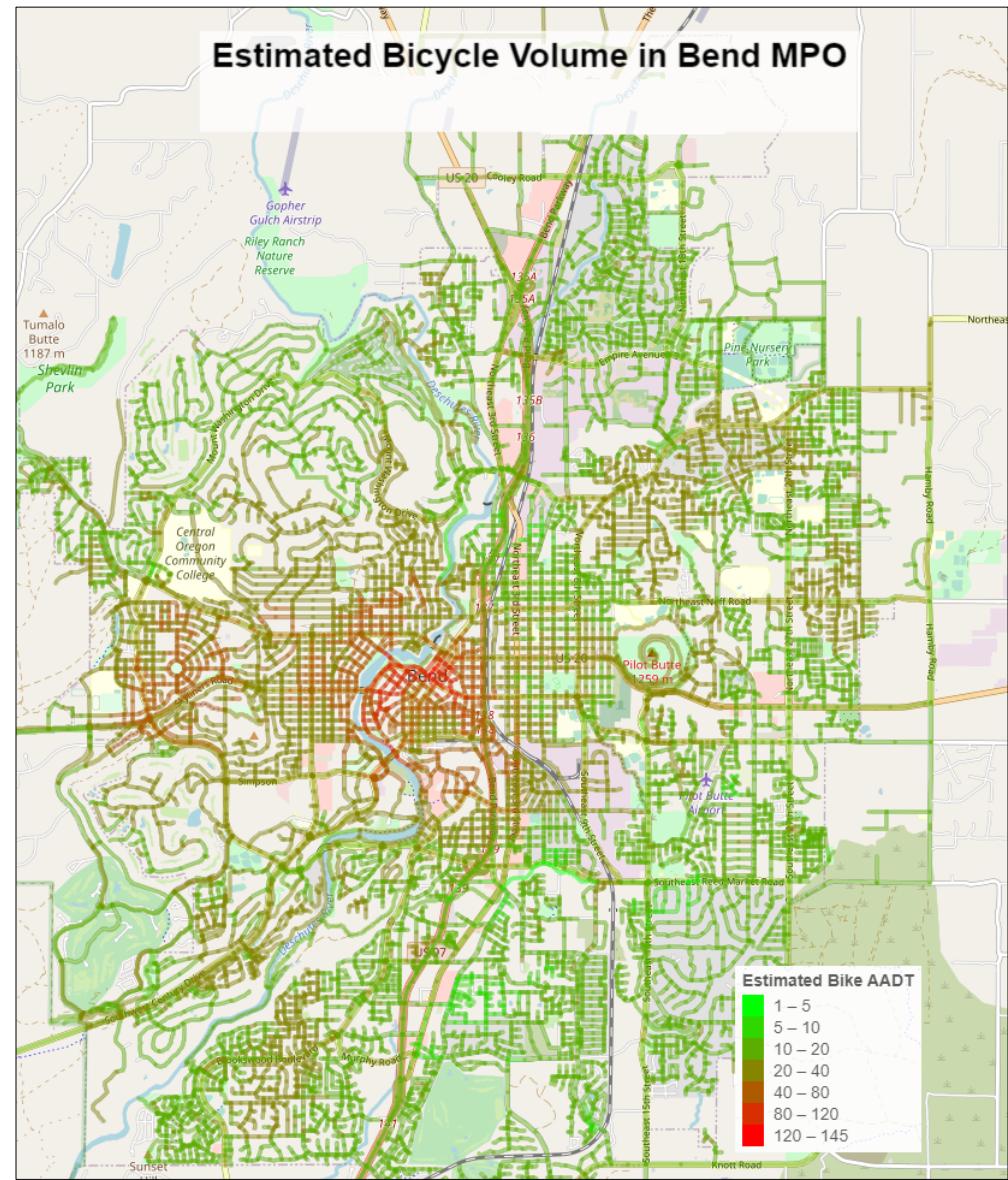
- 10-fold Cross-validation
 - Multiple specifications tried – without Strava and with
 - Overall 43% error (All + Strava model)
 - Prediction error varies by volume bin
 - Low Volumes makes modeling a challenge
 - Probe data helps in accuracy (but even more in application)





Bicycle AADT Model Results

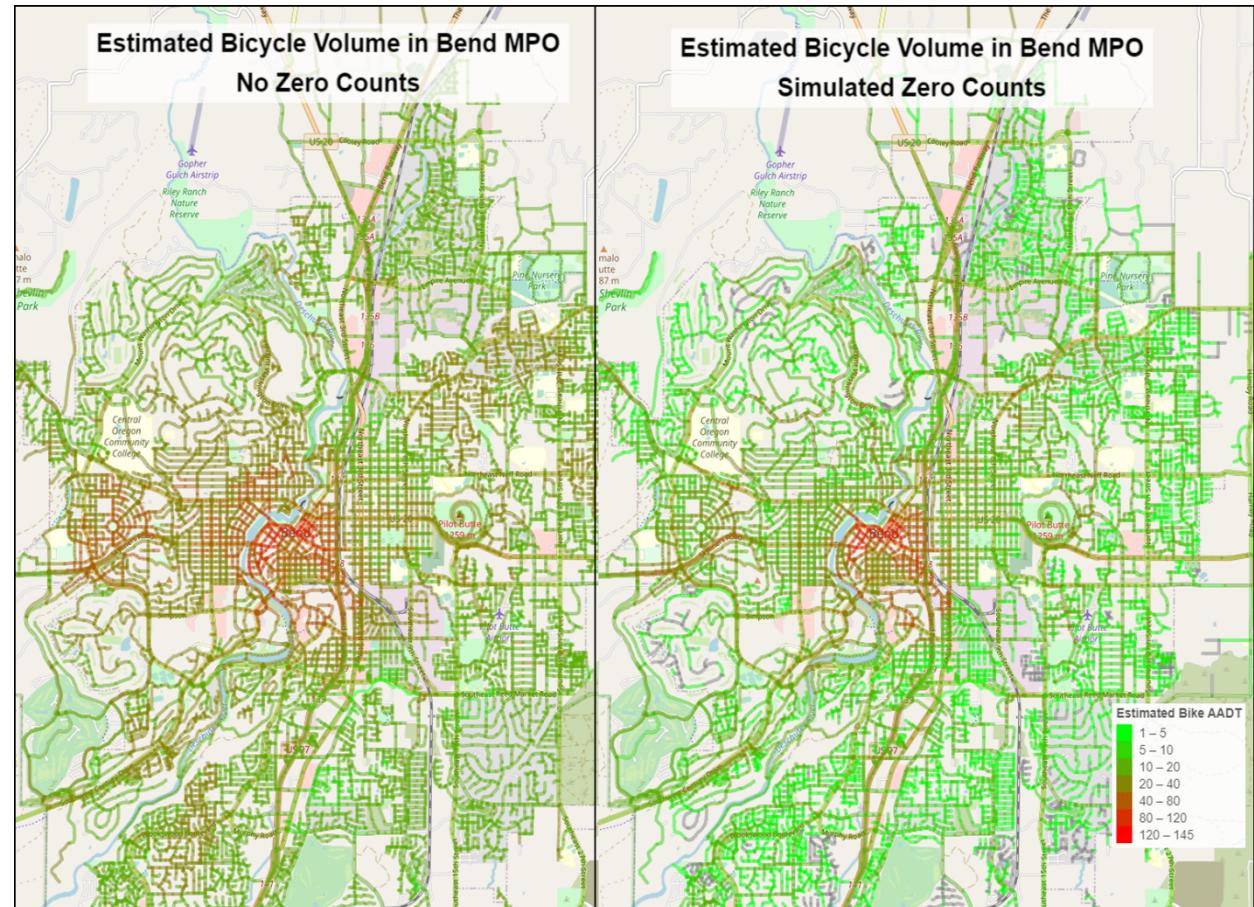
- Network wide estimates
 - Looks reasonable, but how to tell?
 - Activity concentrated near employment centers
 - Appears to estimate too much bike activity in low density residential areas
 - Handling Lack of Zero Counts
 - Random selection of streets high likelihood of zero bike traffic
 - Criteria: local street; low population; density; low centrality, no Strava, no bike facility





Bicycle AADT Model Results

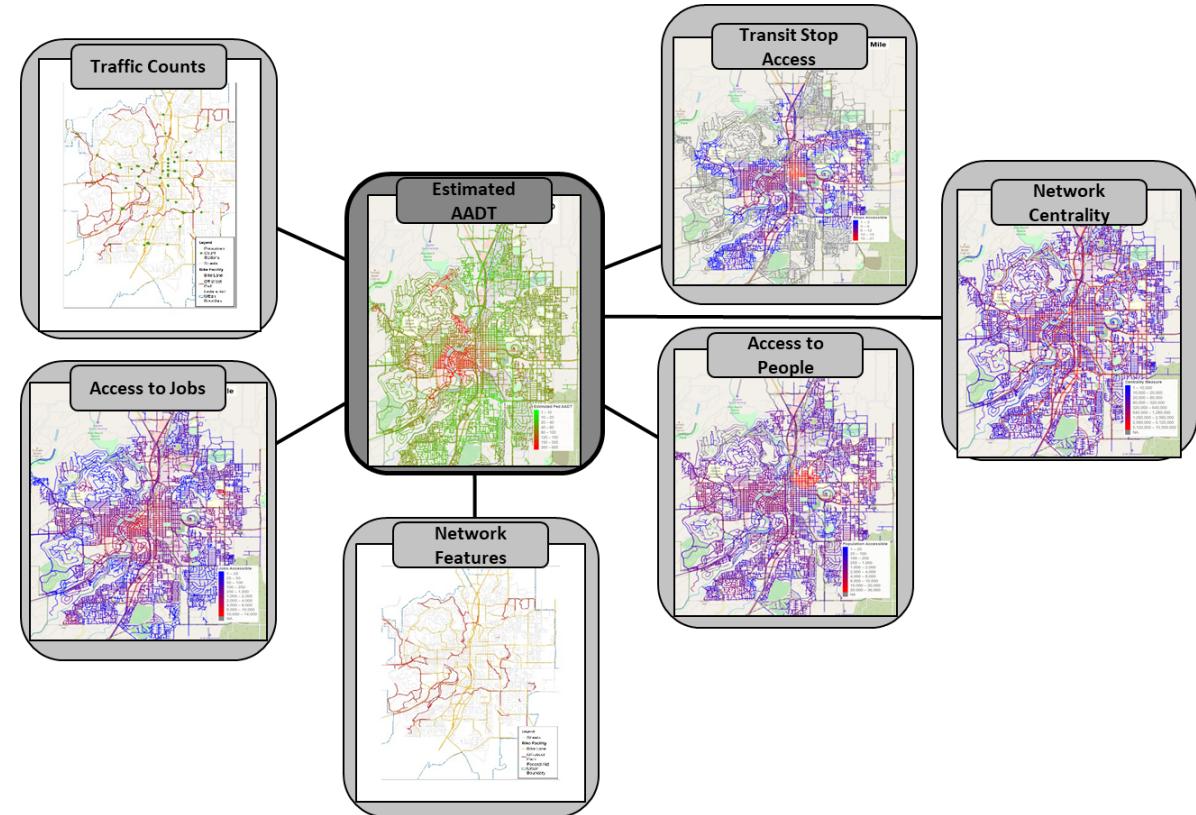
- Handling Lack of Zero Counts
 - Random selection of streets high likelihood of zero bike traffic
 - Criteria: local street; low population; density; low centrality, no Strava, no bike facility
- Results
 - Moderates volume well in expected areas
 - Significantly decreases overall BMT





Pedestrian AADT Data Fusion Scheme

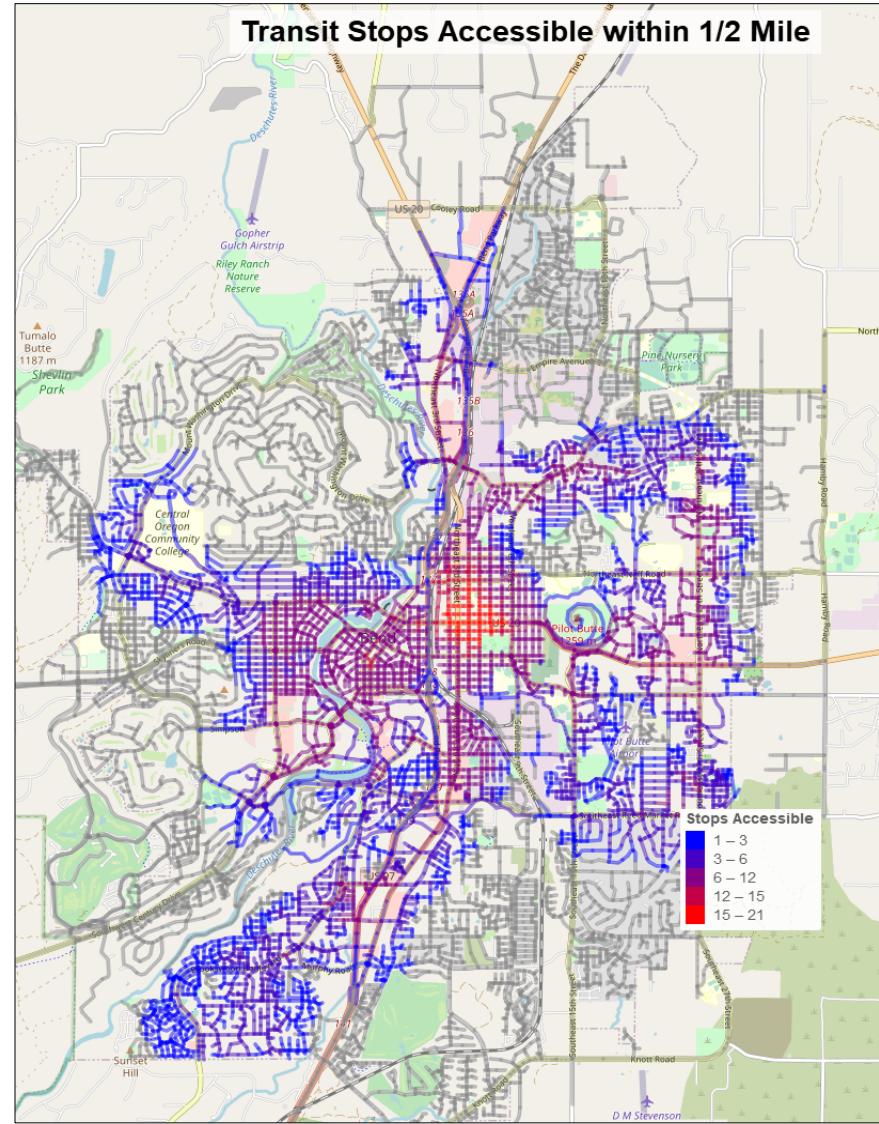
- Pedestrian Model Objectives
 - Provide network wide estimates of pedestrian traffic
- Data and Models Used
 - Up to 512 data features in some specs
 - XgbBoost & Random Forest
 - Census, TAZ, properly attributed routable network, and transit data
- Validation
 - Internal 10-fold cross validations (random partitions)
 - External 10-fold (stratified partition)
 - Leave-one-out validation





Pedestrian AADT Model Data

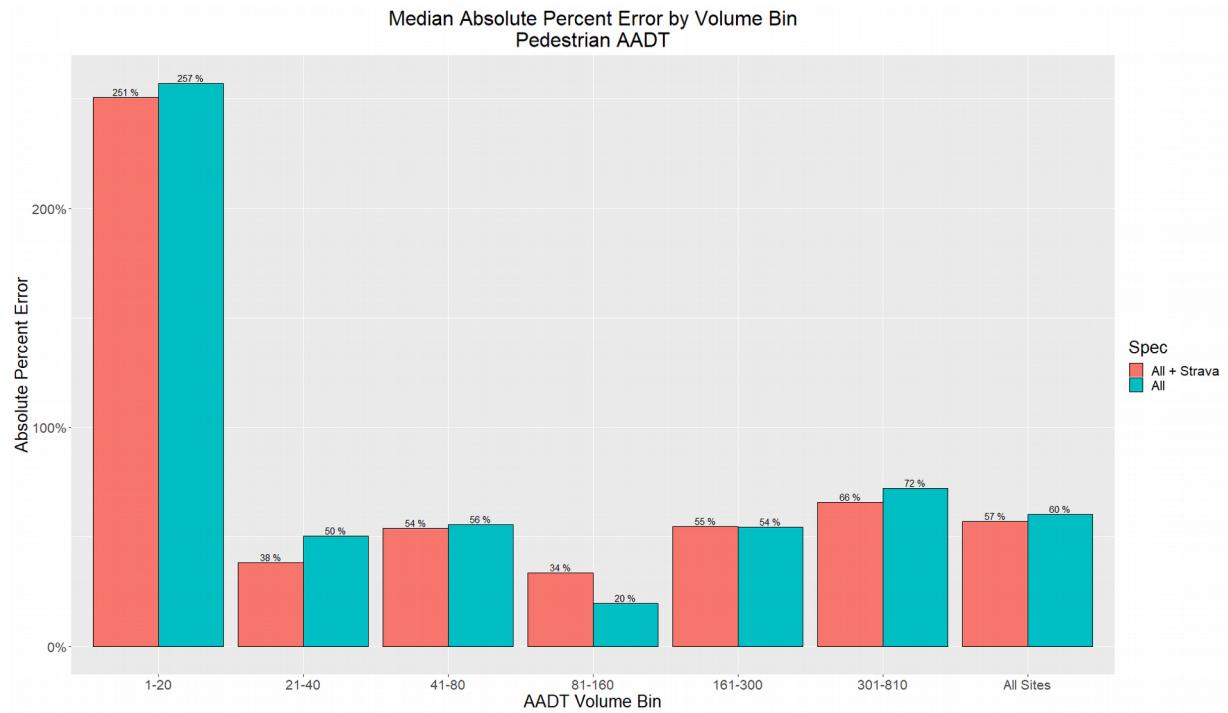
- Bicycle Model Data
 - Traffic Counts
 - 2017, 2018 & 2019
 - N = 56
 - Network Features
 - Functional classification
 - Posted speed limit
 - Off street system
 - Centrality
 - Commute
 - Recreational
 - Shortest
 - Accessibility (distance)
 - Jobs
 - People
 - Transit Stop Access
 - Ridership would be better





Pedestrian AADT Model Validation

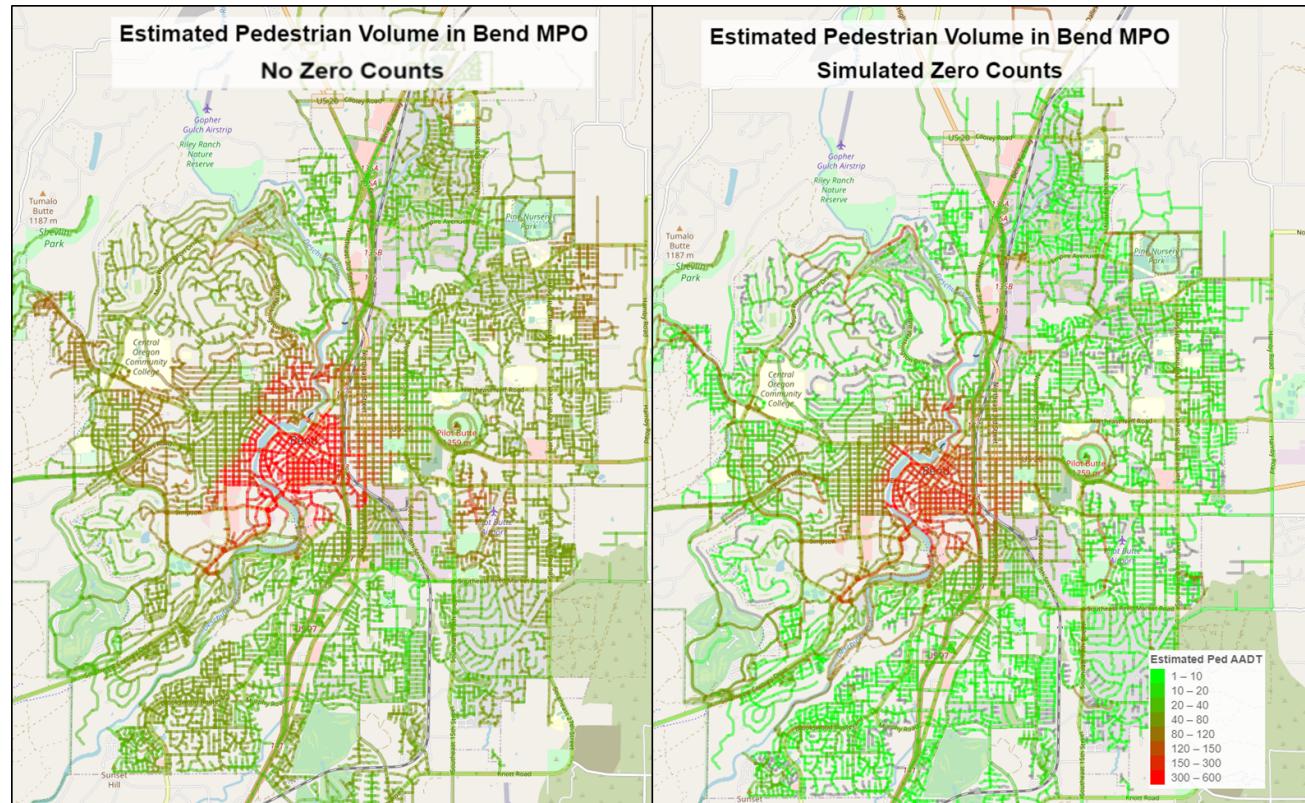
- 10-fold Cross-validation
 - Multiple specifications tried – without Strava and with
 - Overall 57% error
 - Prediction error varies by volume bin
 - Low volumes makes modeling a challenge
 - Probe data helps (surprisingly)





Pedestrian AADT Model Results

- Network wide estimates
 - Looks reasonable, but how to tell?
 - Activity concentrated near employment centers
 - Appears to estimate too much bike activity in low density residential areas
- Handling Lack of Zero Counts
 - Random selection of streets high likelihood of zero bike traffic
 - Criteria: local street; low population; density; low centrality, no Strava, no bike facility



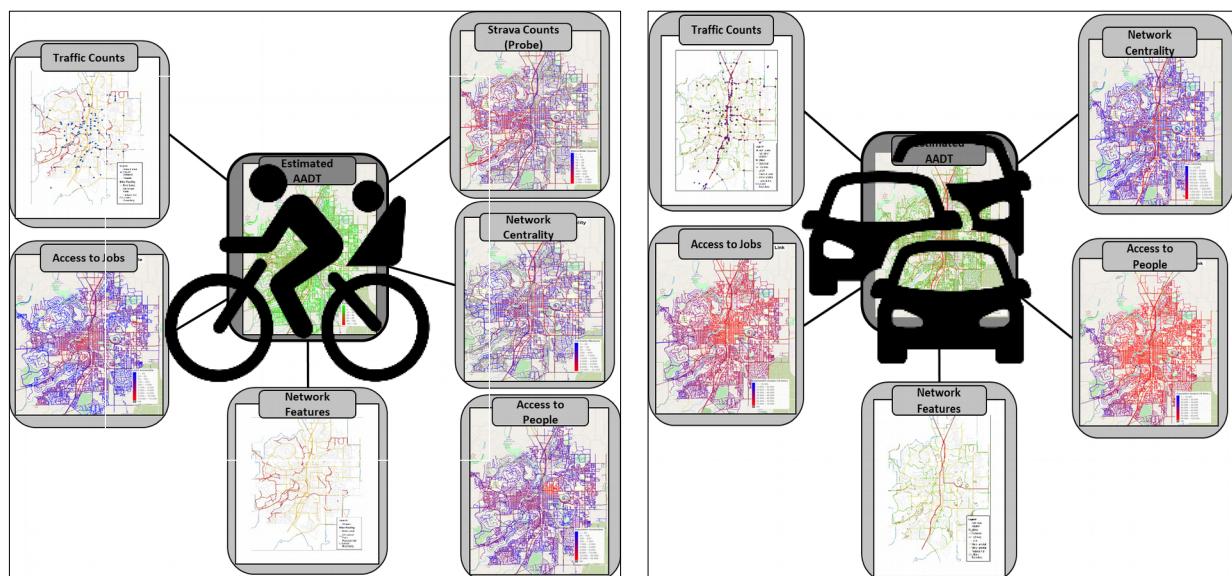
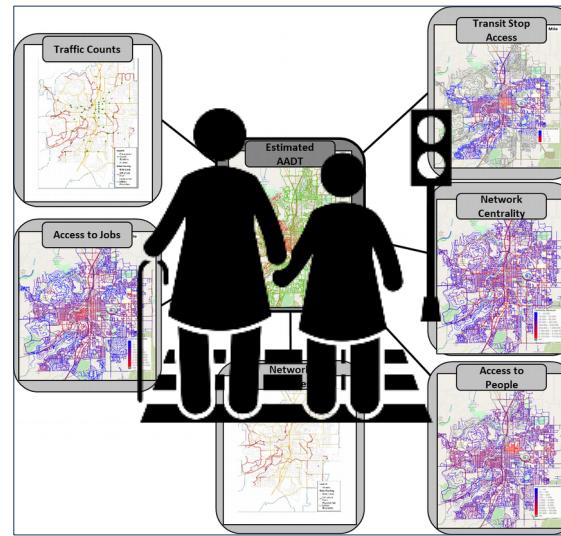
Data Fusion Wrap-up

Limitations

- Need more counts
- Input features not all concurrent with counts (population & employment)
- No probe data for vehicles (or ped specific)
- Feature space could be reduced

Conclusions

- Information from models can inform multiple purposes
- More counts will improve the model
- Future discussions needed to determine further applications



Research Project Next Steps

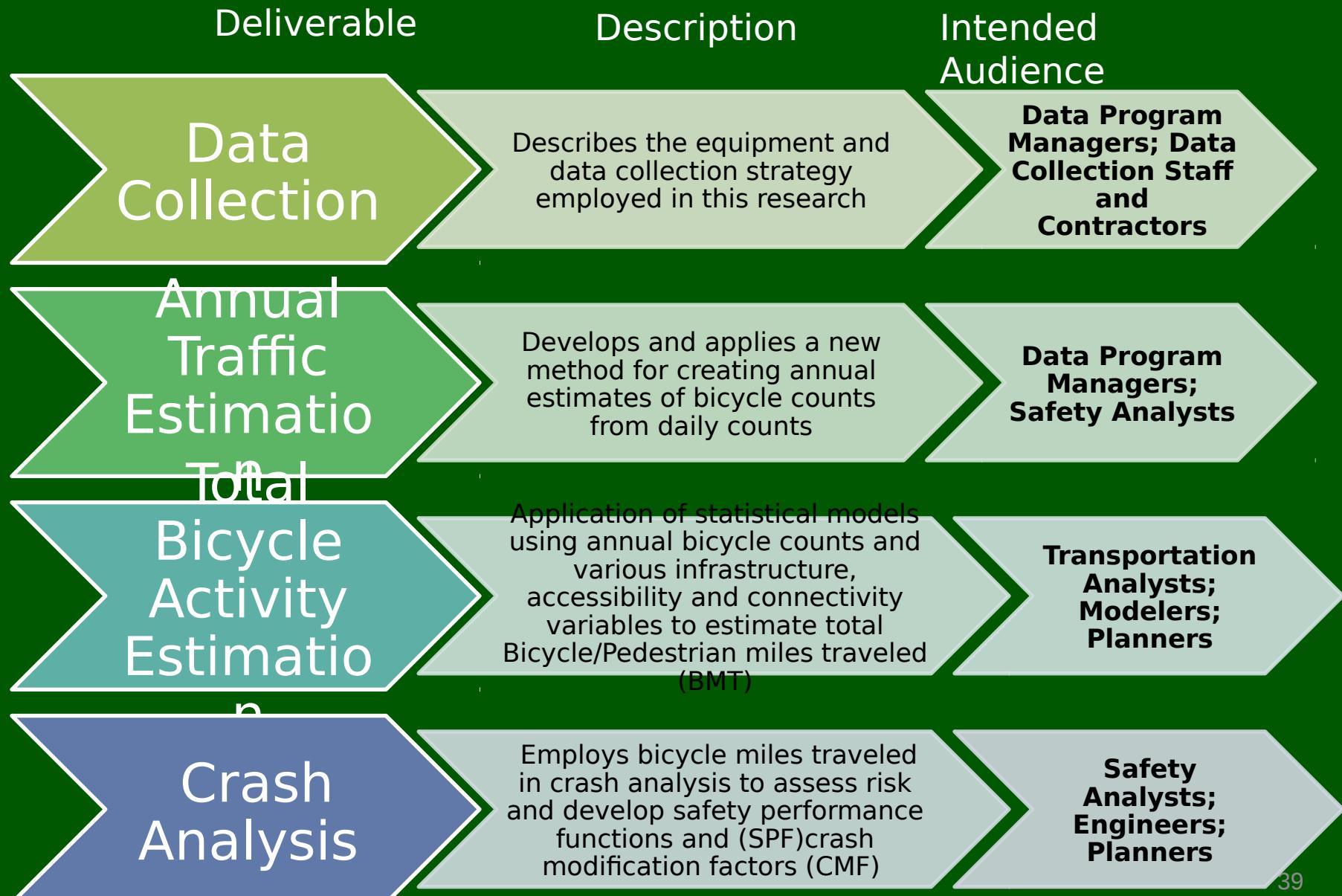
Short term

- Crash data analysis – TAC meeting # 4
- Transfer data processing and related knowledge to Bend area staff
- Provide ongoing tech support for CPiR
- Develop useful data visualizations and data access
- NITC Pooled Fund December 2020

Longer term

- Statewide data support (centralized repository, QAQC) – many pathways to statewide program
- Institutionalize data fusion models for monitoring planning (incorporate NITC results)
- Pilot in another Oregon urban area
- UMD and I-95 Corridor Coalition (RITIS?)
- Better prepare for third-party platform offers – more evaluations of products (e.g. Streetlight Data evaluation)

Final Report (June/July 2020)



National Institute of Transportation & Communities Pooled Fund

Objective

- Develop acceptance criteria for 3rd party data
- Activity estimates for entire network (just bikes)

Partners

- Oregon (Bend MPO, Central Lane MPO, PBOT, ODOT)
- Colorado DOT
- Virginia DOT
- Utah DOT
- DCDOT

Exploring Data Fusion Techniques to Derive Bicycle Volumes on a Network



Sirisha Kothuri
Joe Broach
Nathan McNeil

 Portland State
UNIVERSITY

Kate Hyun
Steve Mattingly

 UNIVERSITY OF
TEXAS
ARLINGTON

Krista Nordback

 THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Frank Proulx

 TOOLE
DESIGN



Questions



Oregon Department of Transportation

Questions?

Josh Roll Active and
Sustainable Transportation
Research Coordinator
Josh.F.Roll@ODOT.state.or.us



Back up



Traffic Data Imputation - What's the Problem?

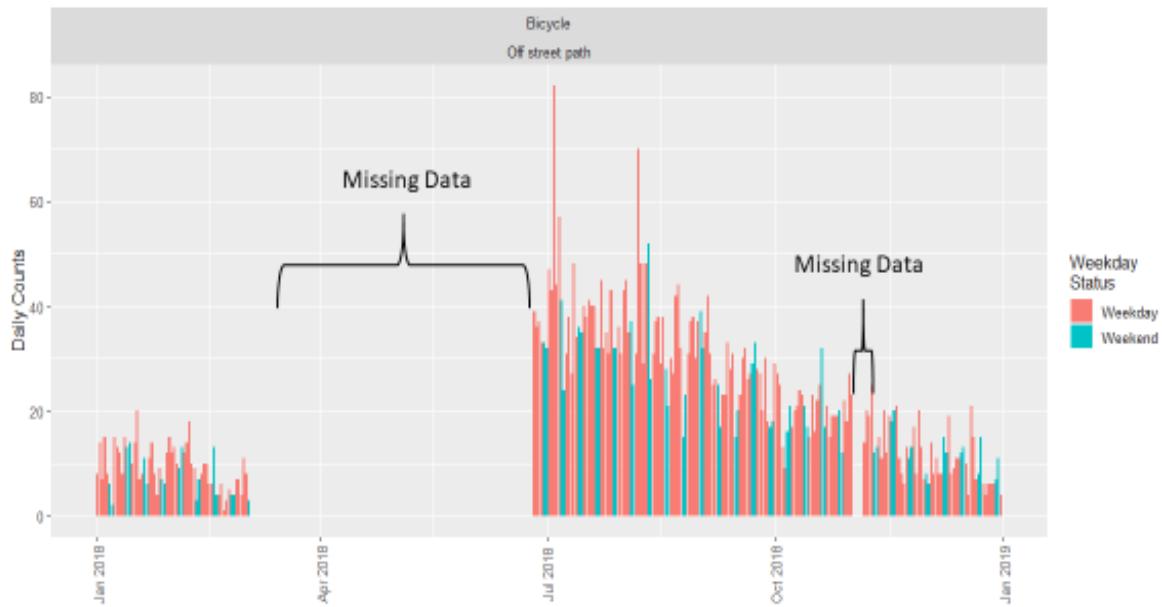
Why Missing Data?

- Equipment failure at permanent sites (bugs!)
- Data Transfer Issues

Franklin Undercrossing WB Multilane Path west of PED tunnel under US97 Parkway and Rail

Solution:

- Traffic variation highly dependent on weather and day of week factors
- (Hanson and Hanson 1977; Niemeier 1996; Nankervis 1999; Richardson 2000; Brandenburg 2007; Rose et al. 2011; Tin Tin et al. 2012, Thomas, Jaarsma, and Tutert 2009; Lewis 2011; Gallop, Tse, and Zhao 2012; Miranda-Moreno and Nosal 2011; Nosal and Miranda-Moreno 2012; Schmiedeskamp and Zhao 2016).



Daily Imputation and Annual Estimation

Results by Months Used

- More months of data equals better results
- Likely scenario is 3 months or less of missing data
- 2-10% error when 9 months of data used

Limitations

- Only using 1 year of data but results would be better if multiple years of data are used
- Negative Binomial does poorly when data poor

Number of Months Used in Training	Bicycle					Pedestrian				
	Negative Binomial		Random Forest			Negative Binomial		Random Forest		
	95th Pct.	Median	95th Pct.	Median	95th Pct.	Median	95th Pct.	Median	95th Pct.	Median
1	26,288%	38%	84%	34%	244%	18%	68%	20%		
2	131%	14%	53%	11%	53%	9%	40%	10%		
3	50%	10%	34%	7%	28%	6%	27%	7%		
4	35%	7%	23%	5%	19%	5%	20%	6%		
5	28%	5%	18%	4%	14%	4%	17%	4%		
6	22%	4%	15%	3%	11%	3%	14%	4%		
7	18%	3%	12%	2%	9%	2%	12%	3%		
8	15%	3%	10%	2%	7%	2%	10%	3%		
9	12%	2%	8%	1%	6%	2%	8%	2%		
10	9%	2%	6%	1%	4%	1%	6%	2%		
11	6%	1%	4%	1%	3%	1%	4%	1%		



Wait what is Machine Learning Again?

Node
Key

Tree Split Stopping Rules/Criteria

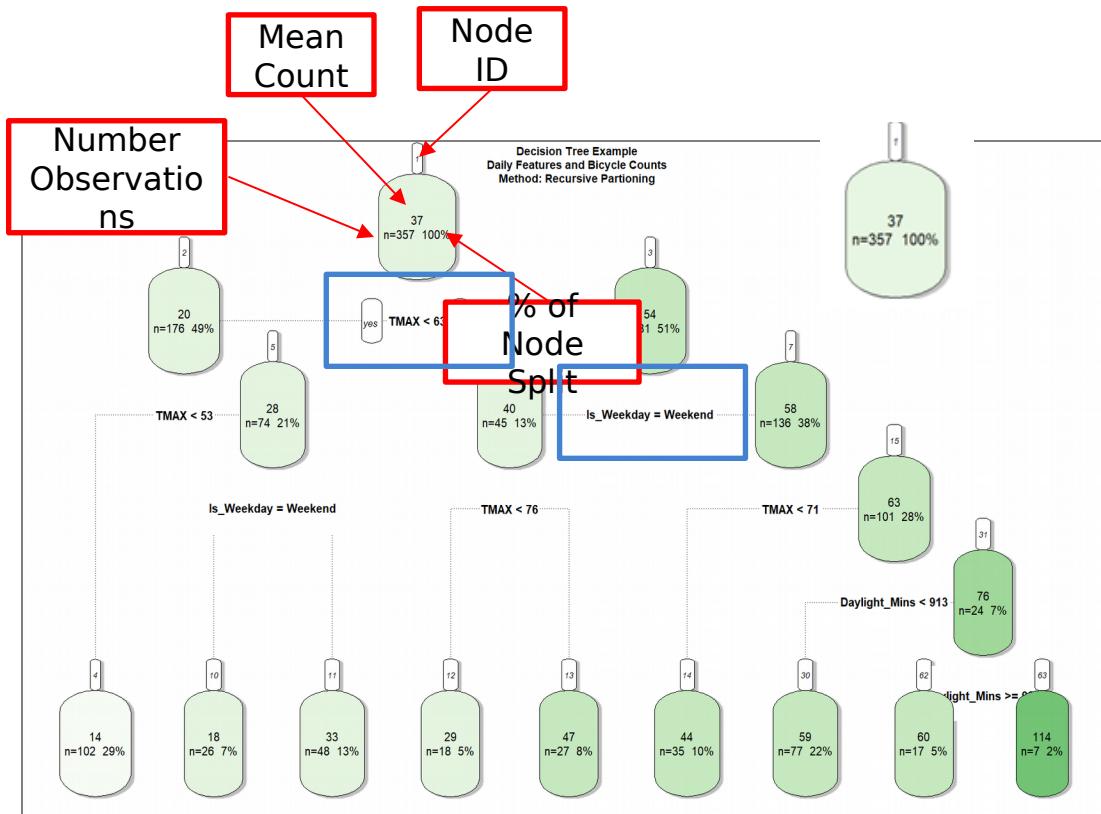
- Guided by rules of impurity reduction with an aim of creating daughter nodes more pure than parent nodes
- Impurity quantified by GINI Index or Shannon Entropy
- Given a minimum # of observations left in node

Traffic Count Imputation Example

- TMAX – most important
- Weekday variable – also important
- Minutes of daylight – *also* important

Ensembles

- Example is single tree
- Multiple trees estimated
- Combined to create a forest!



Daily Imputation and Annual Estimation

Data

- 21 unique locations from statewide data
- All sites have at least 98% of annual data

Imputation

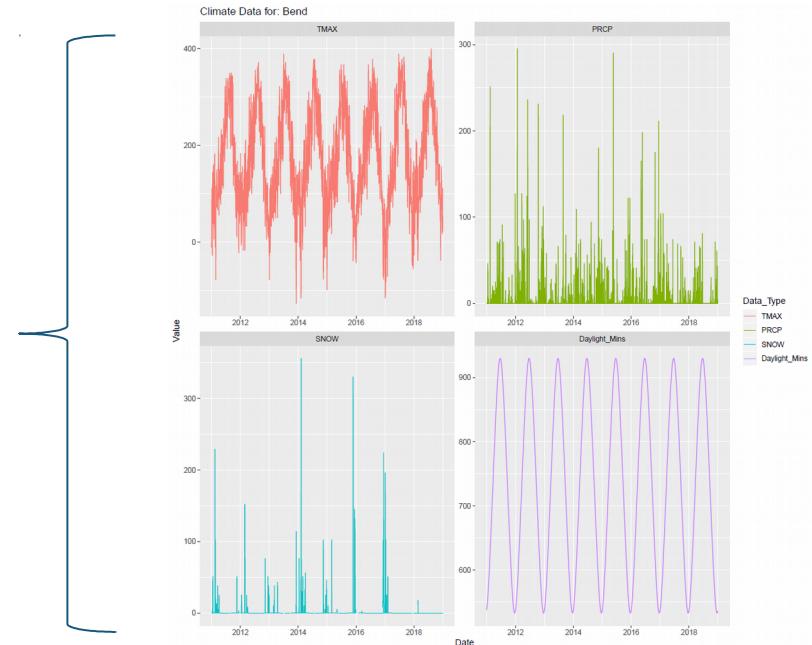
- Machine learning (random forest, conditional inference, recursive partitioning)
- Negative binomial regression

Test setup

- Use permanent counters from around the state
- Hold out all possible combinations of month

City	User Type	Daily Counts Summary				Number Locations	
		Mean	Median	Std. Dev.	Records	Unique	Year/Location*
Bend	Bicycle	56	43	55	2,167	5	6
Bend	Pedestrian	148	99	150	2,907	7	8
Eugene	Bicycle	340	275	240	1,095	3	3
Eugene	Pedestrian	491	303	450	1,824	5	5
Portland	Bicycle	1,957	1,720	1,402	728	1	2
Salem	Bicycle	38	32	32	365	1	1
Springfield	Bicycle	185	125	182	1,460	4	4
Springfield	Pedestrian	103	97	42	365	1	1
Total	Total	327	116	627	10,911	21	30

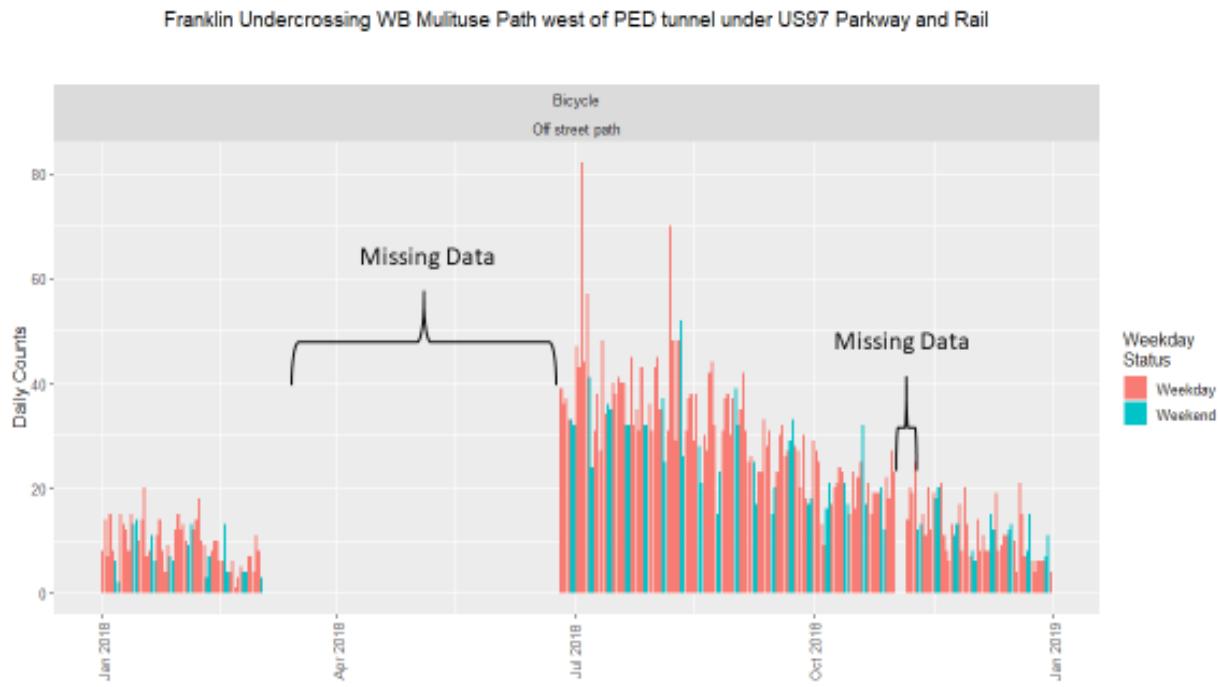
NOAA Data



Why Machine Learning?

Why Machine Learning?

- Negative Binomial Regression used previously (SARM) Roll and Proulx 2017
- Shown to predict annual traffic within 5% with just 3 weeks of counts
- But how to select best model?
- Interaction effects better captured in ML



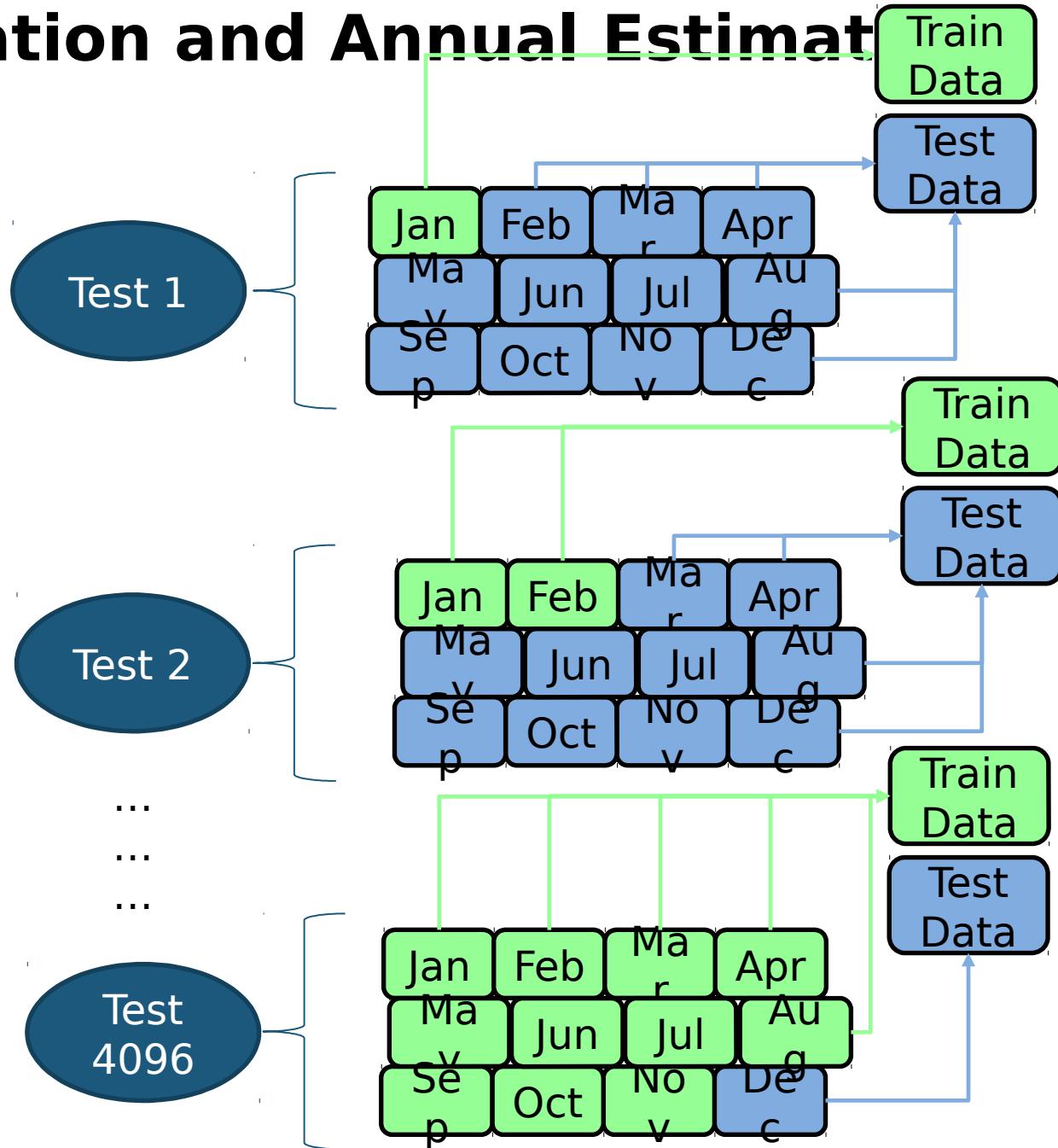
Daily Imputation and Annual Estimation

Imputation

- Machine learning (using recursive partitioning regression trees, random forests, conditional inference)
- Negative binomial regression

Test setup

- Use permanent counters from around the state
- Estimate **daily** traffic counts
- Hold out all possible (4,096) combinations of month
- Measure monthly and annual error



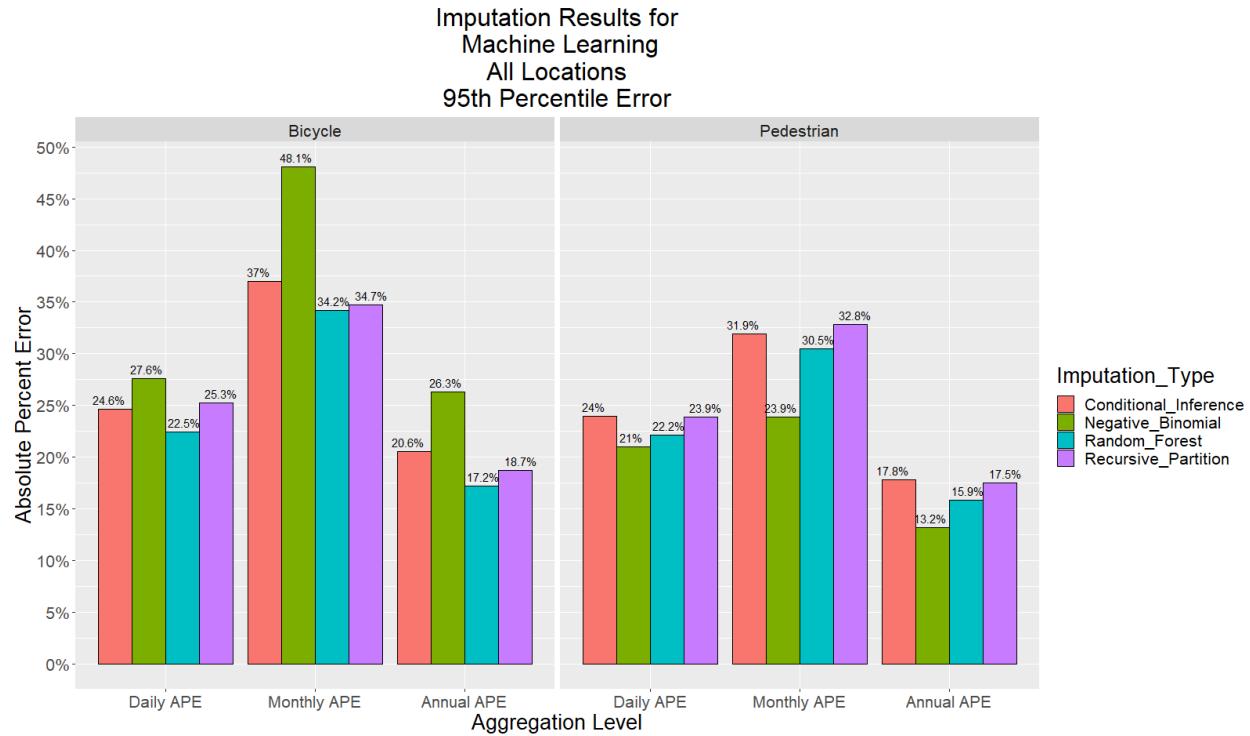
Daily Imputation and Annual Estimation

Results

- 3 levels of estimation
- Bikes – Random Forest works best
- Peds – Close tie between negative binomial and random forest

Test setup

- Use permanent counters from around the state
- Hold out all possible combinations of month



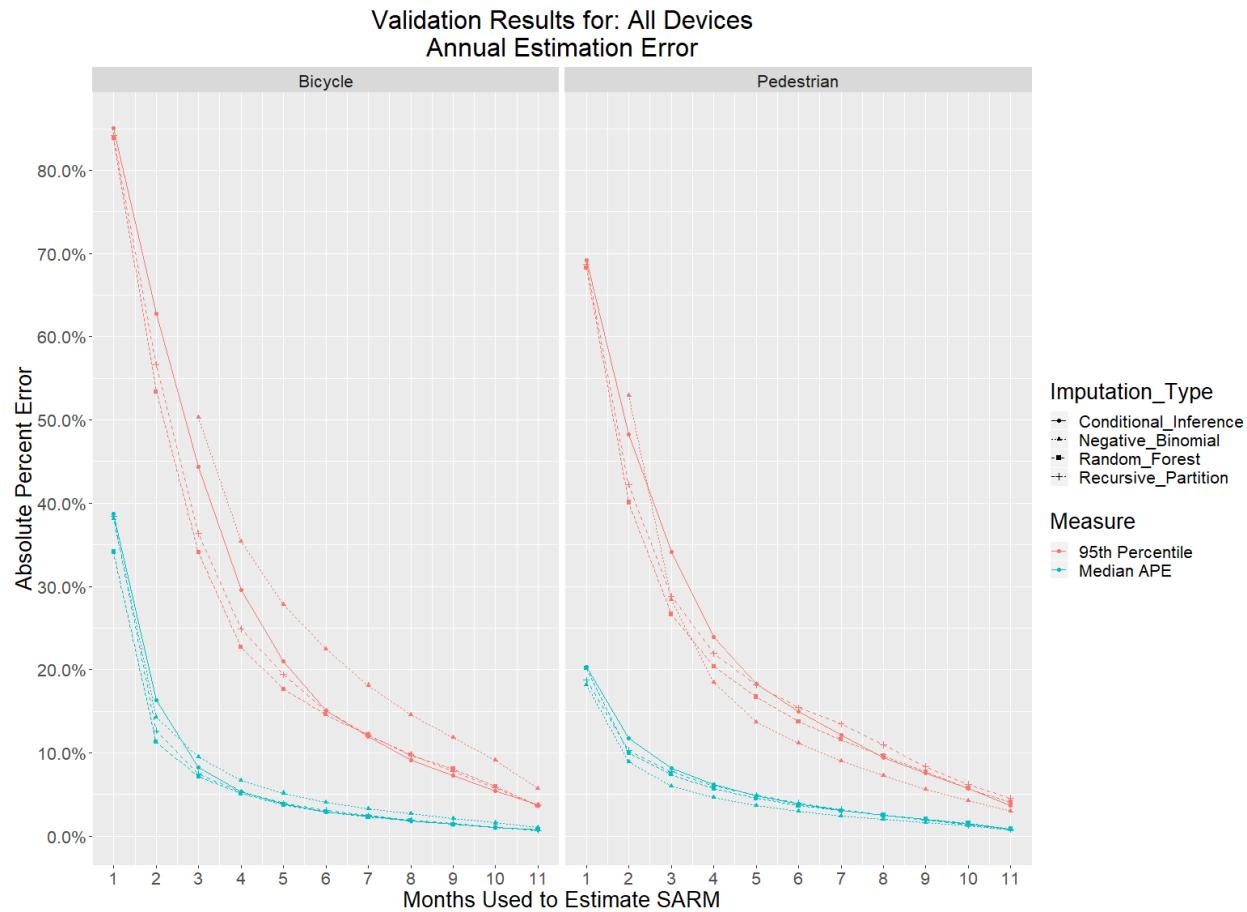
Daily Imputation and Annual Estimation

Results by Months Used

- More months of data equals better results
- Likely scenario is 3 months or less of missing data
- 2-10% error when 9 months of data used

Limitations

- Only using 1 year of data but results would be better if multiple years of data are used
- Negative Binomial does poorly when data poor



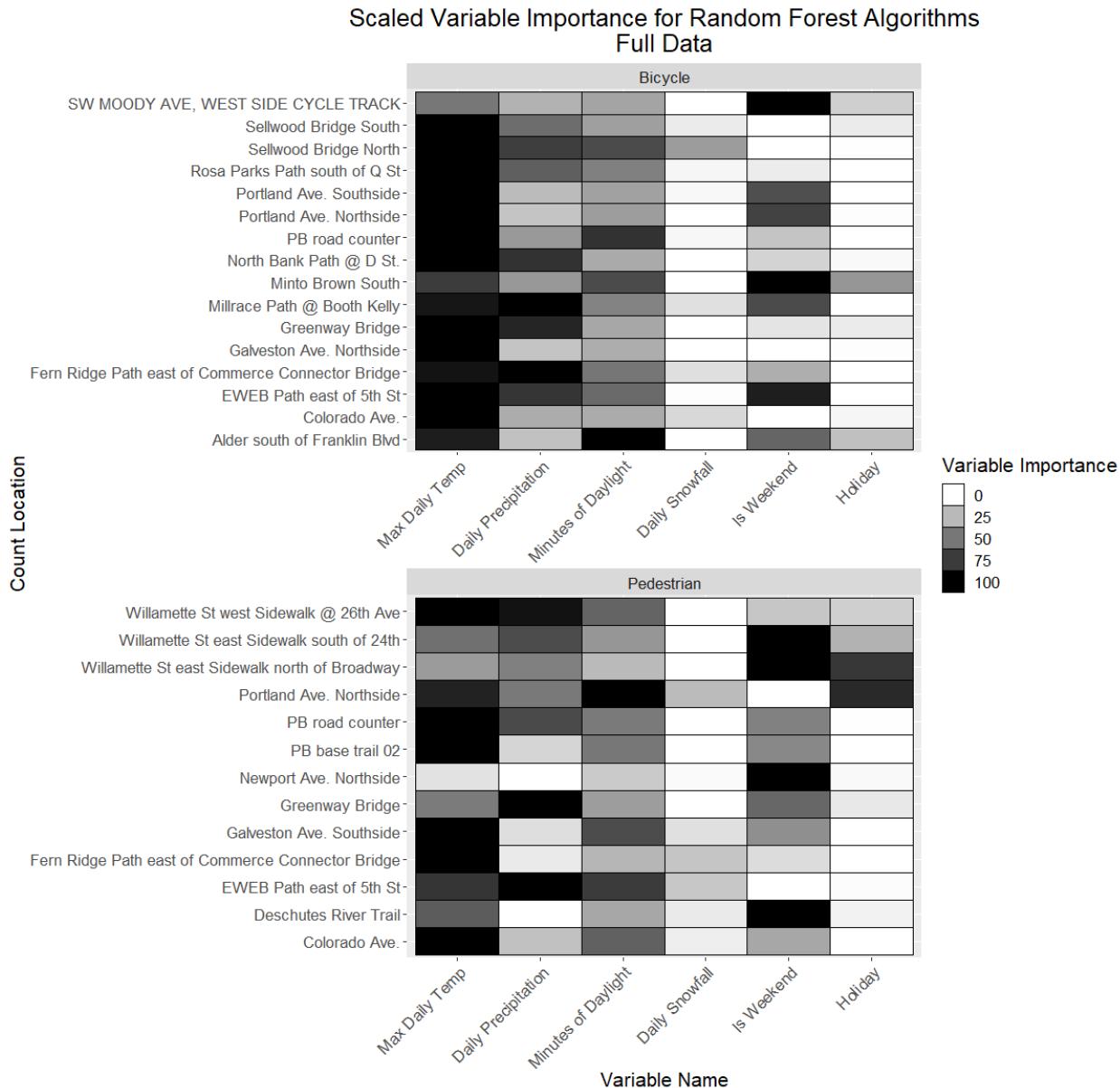
Variable Importance

What Variables Are Important in ML Algorithm?

- Inference generally a limitation of ML
- But variable importance can be calculated (at computational cost)
- Measure of node purity

Variable Importance Results

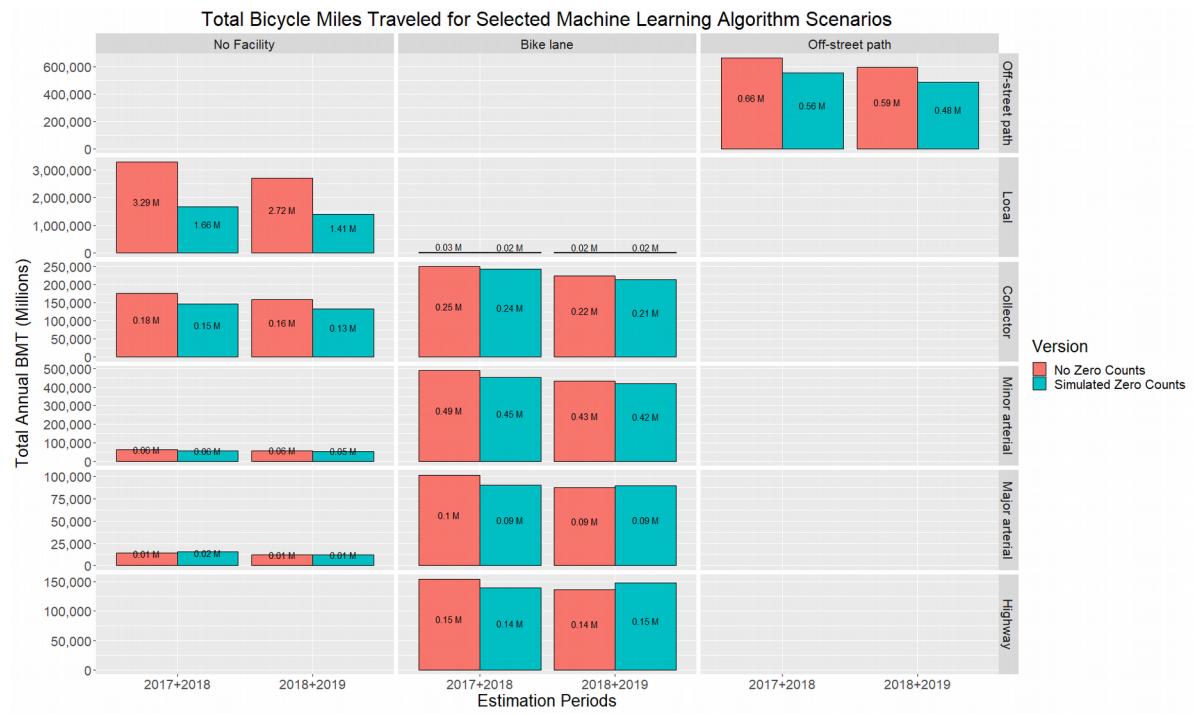
- Temperature importance in all models
- Precipitation and daylight next most important





Bicycle AADT Model Results

- Handling Lack of Zero Counts
 - Random selection of streets high likelihood of zero bike traffic
 - Criteria: local street; low population; density; low centrality, no Strava, no bike facility
- Results
 - Moderates volume well in expected areas
 - Decreases overall BMT by about 1/3



Estimation Periods	Total Annual Bicycle Miles Traveled		Percent Difference
	No Zero Counts	Simulated Zero Counts	
2017+2018	5,225,730	3,385,390	65%
2018+2019	4,444,592	2,985,239	67%

