ORCA 2017
One Regional Card for All

Project Lead & Data Scientists:
Mark Hallenbeck, Anat Caspi, Bryna Hazelton,
Michael Wolf, Jake VanderPlas

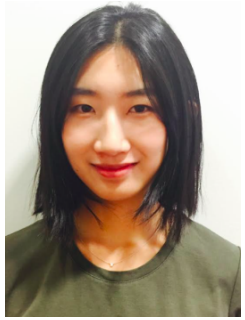
DSSG Fellows

Mayuree Binjolkar

Daniel Dylewsky

Andrew Ju

Wenonah Zhang



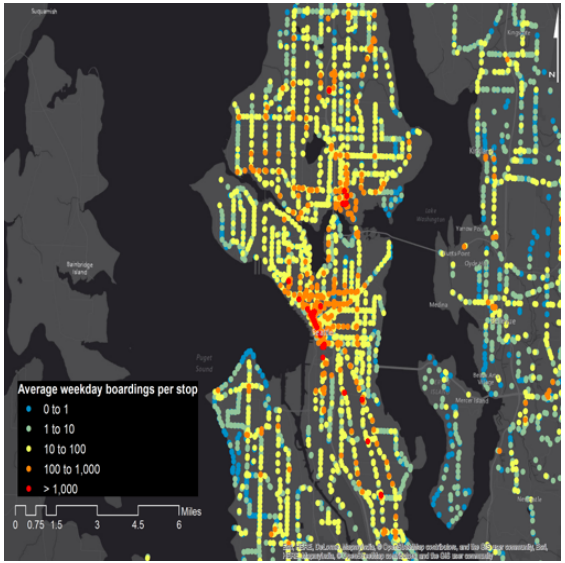
ORCA Transactions Data

Information helps to improve transit performance

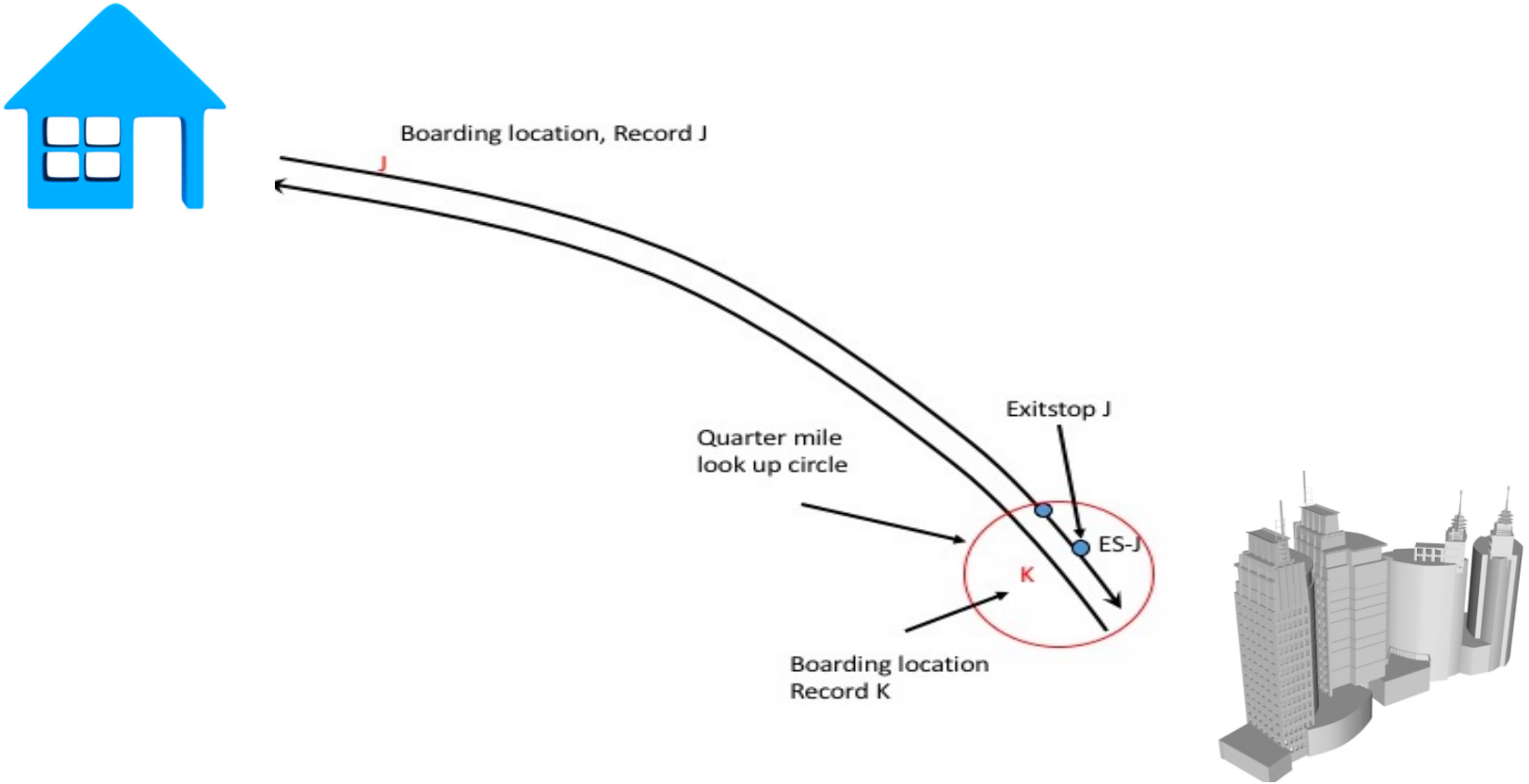
First Step: Geolocate the ORCA boardings

ORCA Transactions Data

Automatic Vehicle Location Data



ORCA Boardings - Only Half of the Story
Estimate Destination of the Trip



Transfer Analysis Objectives

Transfer Data

Real vs. Financial Transfer?

Cleaner Origin/Destination Data

- Headway
- Walking Time
- Transfer Duration
- # bus missed

Model Selection Stage

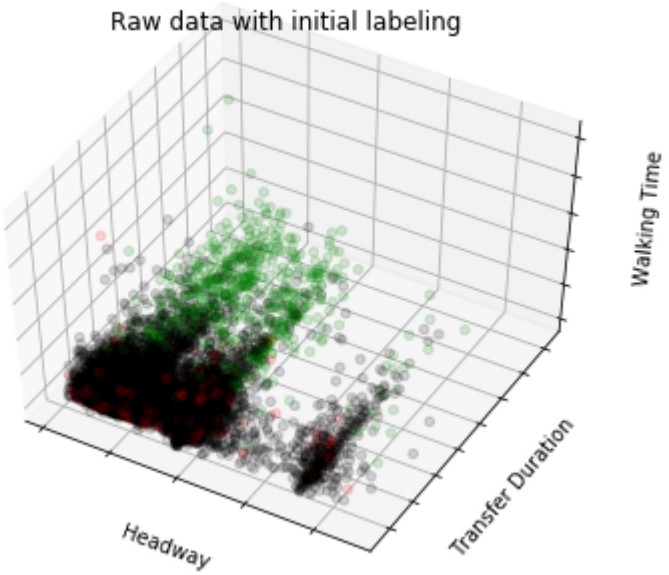
- No ground truth to conduct supervised learning
- Gaussian Mixture Model did not perform as well as expected
- K means Unsupervised learning oversimplified the clusters
- The amount of labeled data based on human intuition is not sufficient for supervised learning

Why Semi-supervised Learning with Label Spreading Algorithm?

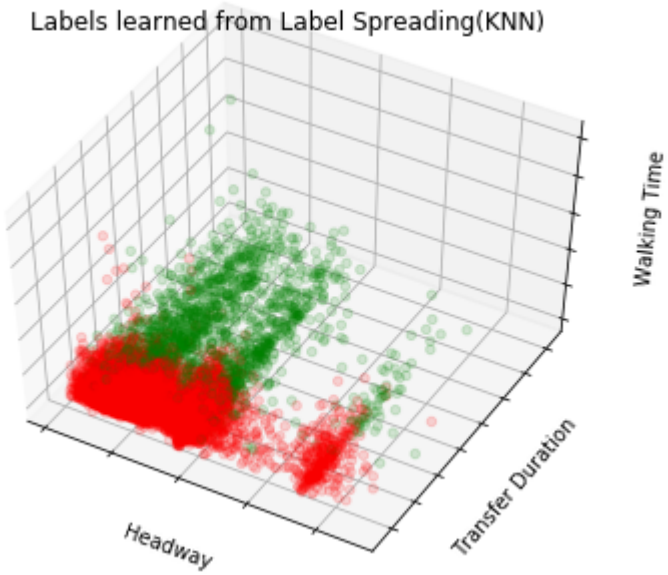
- Performs well with a small amount of labeled data
- Considerable improvement in learning accuracy when use unlabeled data in conjunction with labeled data

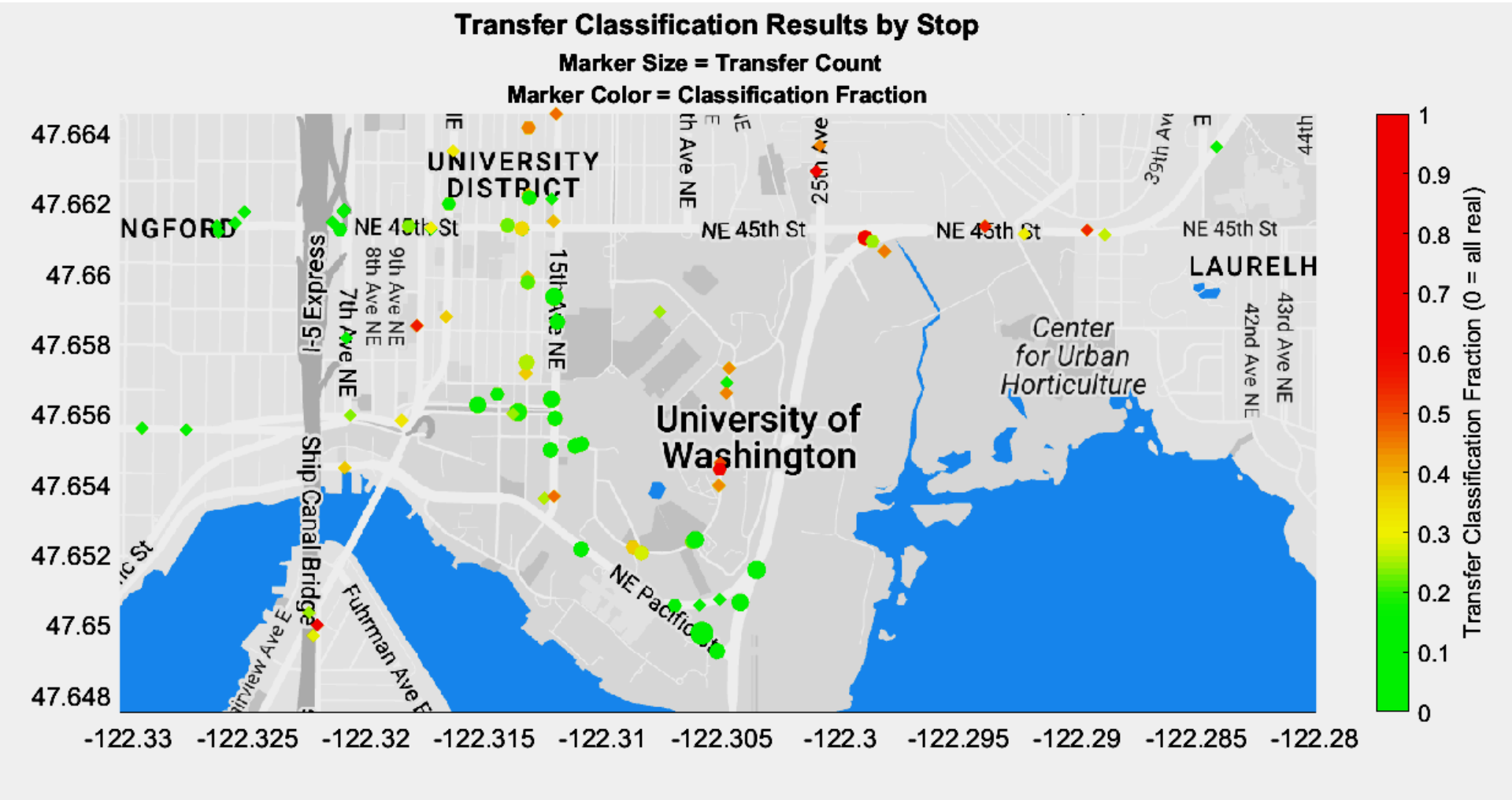
Label Spreading Result

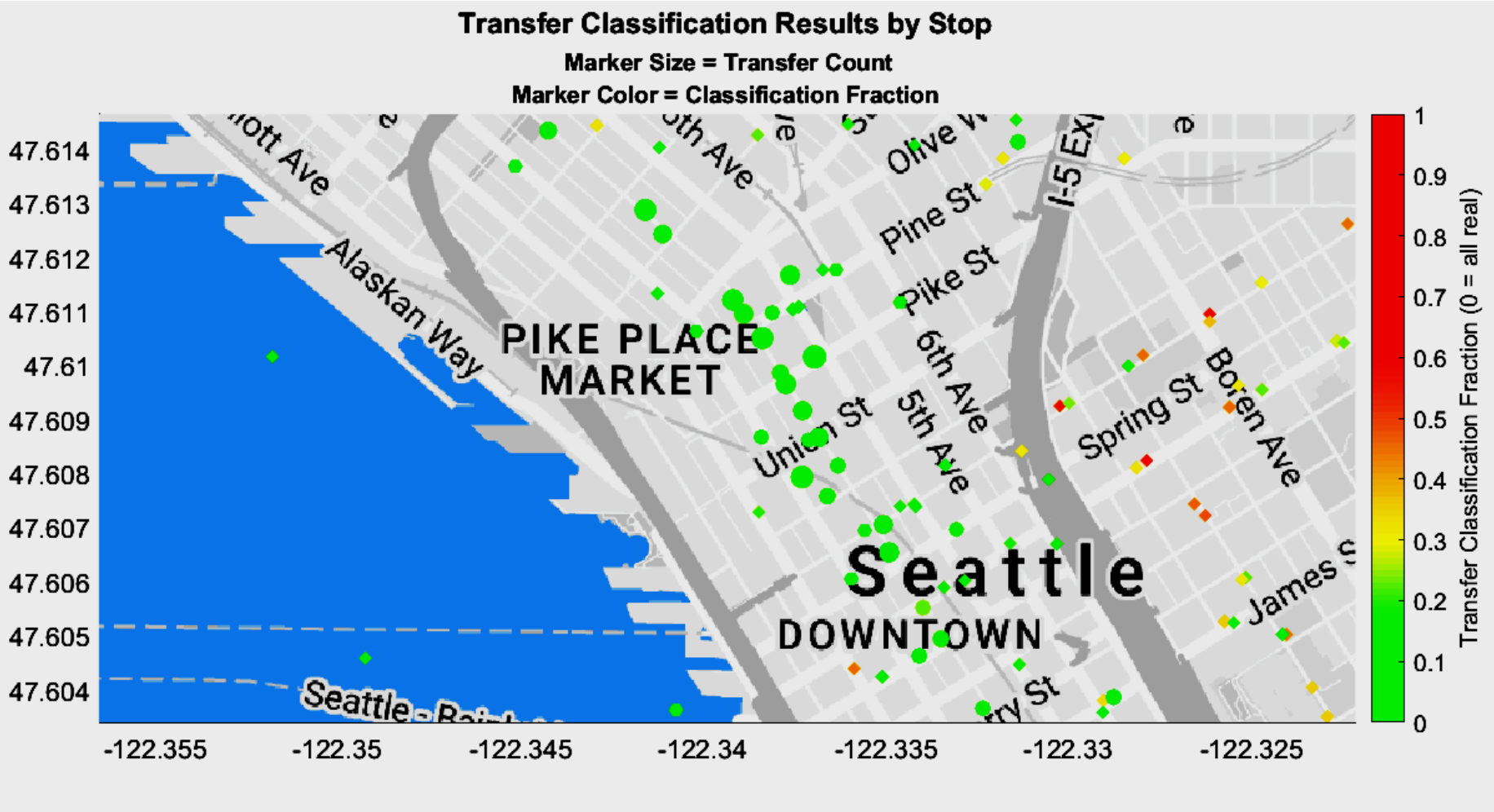
- Real Transfer 5%
- Financial Transfer 7%
- Unlabeled 88%



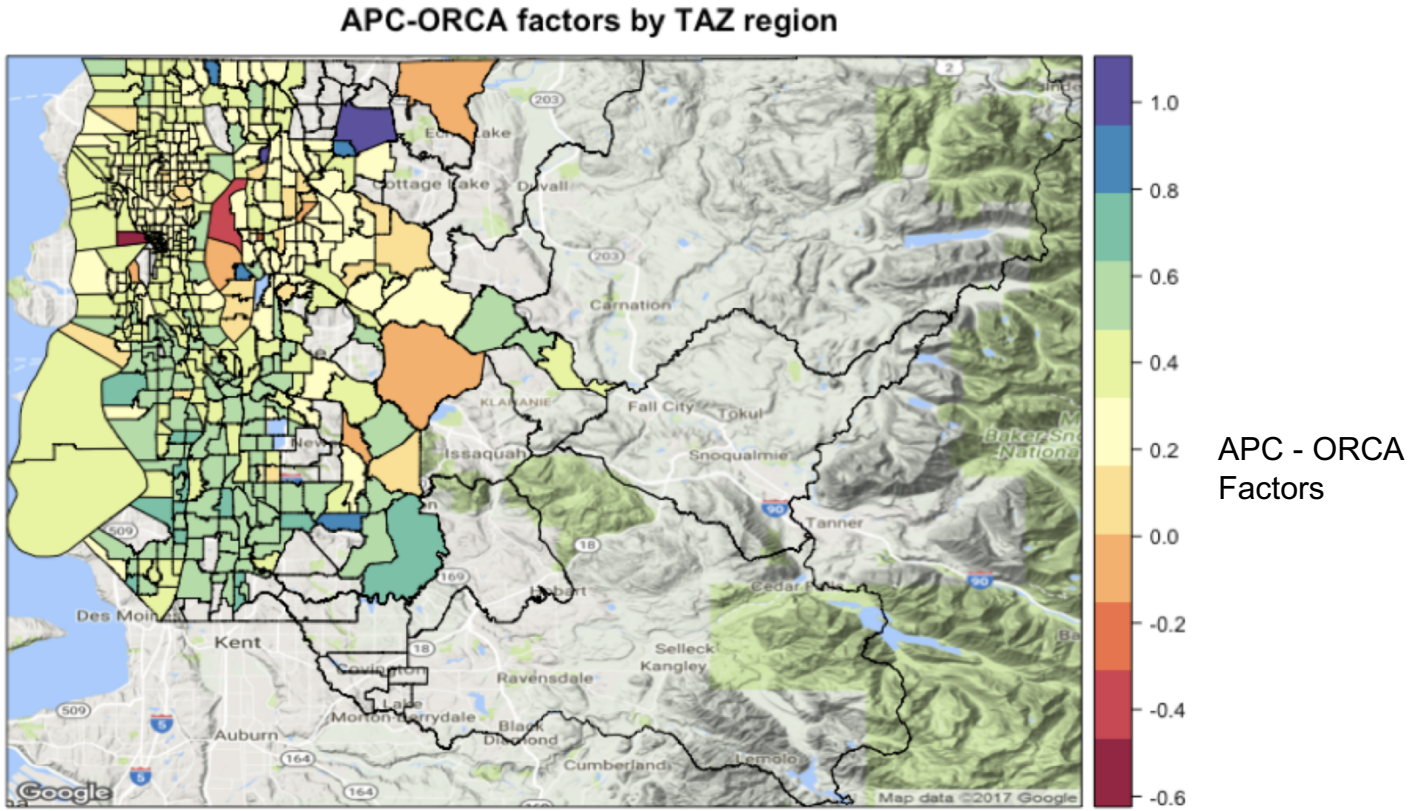
- Real Transfer 83%
- Financial Transfer 17%




Transfer Classification Results by Stop
Marker Size = Transfer Count
Marker Color = Classification Fraction
UNIVERSITY DISTRICT
NE 45th St
University of Washington
Center for Urban Horticulture
LAURELH
I-5 Express
Ship Canal Bridge
Fuhrman Ave E
NE Pacific St
15th Ave NE
9th Ave NE
8th Ave NE
7th Ave NE
25th Ave
42nd Ave NE
43rd Ave NE
47.664
47.662
47.66
47.658
47.656
47.654
47.652
47.65
47.648
-122.33
-122.325
-122.32
-122.315
-122.31
-122.305
-122.3
-122.295
-122.29
-122.285
-122.28
Transfer Classification Fraction (0 = all real)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0


Transfer Classification Results by Stop
Marker Size = Transfer Count
Marker Color = Classification Fraction
PIKE PLACE MARKET
Seattle
DOWNTOWN
Alaskan Way
Pine St
Pike St
6th Ave
5th Ave
Union St
Spring St
Boren Ave
James St
Olive Way
I-5 Exp
47.614
47.613
47.612
47.611
47.61
47.609
47.608
47.607
47.606
47.605
47.604
-122.355
-122.35
-122.345
-122.34
-122.335
-122.33
-122.325
Transfer Classification Fraction (0 = all real)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

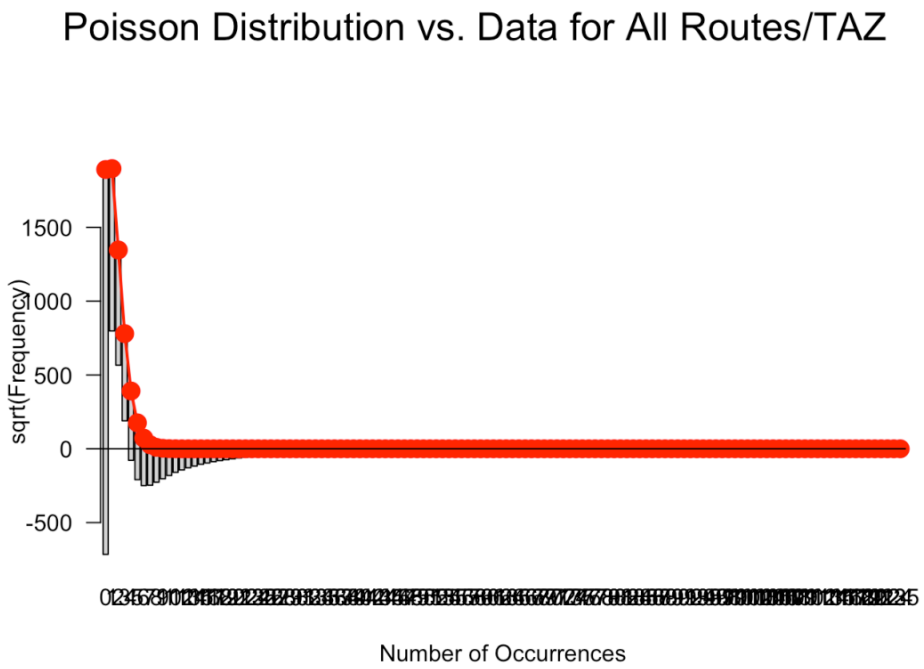
ORCA Data is Biased and Variable



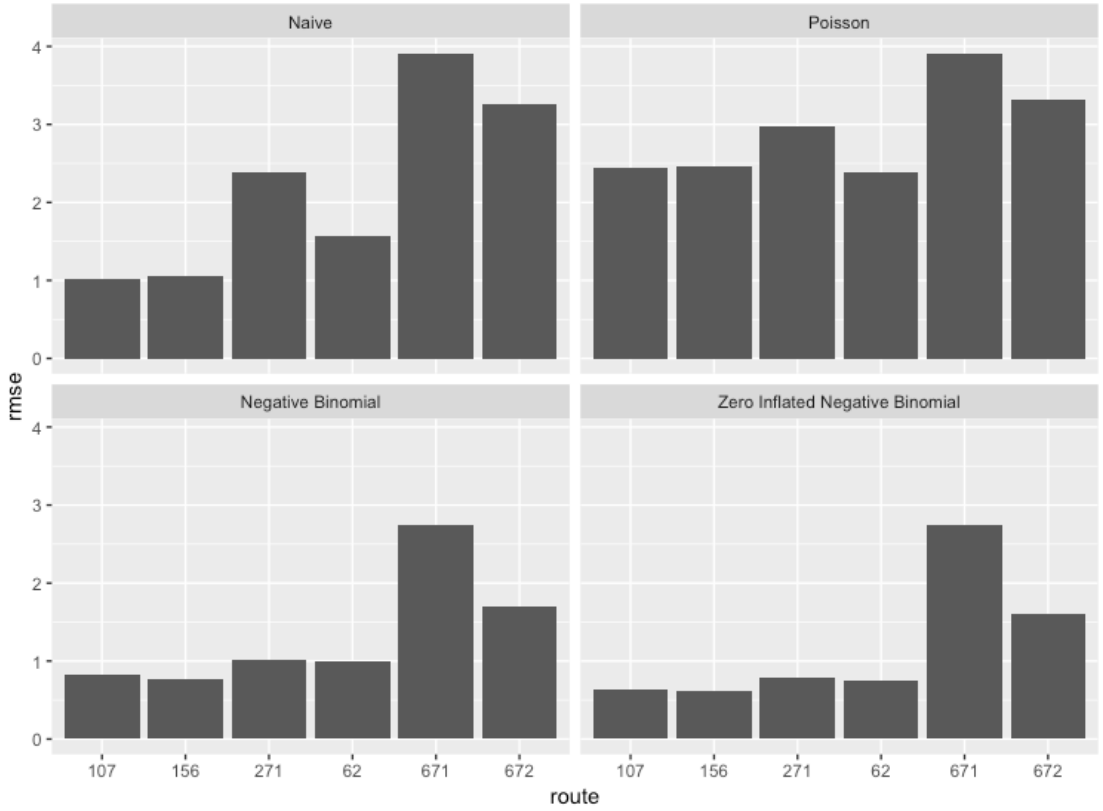
Zero Inflated Negative Binomial Regression

- Regression approach is highly interpretable to understand bias
- Count data is well suited for Poisson distribution
- Overdispersion (mean > variance) suggests a Negative Binomial variant of the Poisson distribution
- High number of 0s and noisy data encourages a zero inflated / mixed model approach

Distribution for APC Count (response variable)



RMSE for Models Across Sample Routes



Continuing Work...

- Validating semi supervised learning models for transfer analysis
- Zero Inflated Negative Binomial Model on Entire Network
- Neural Nets (scalability, diversity)


GORDON AND BETTY
MOORE
FOUNDATION


Washington Research
FOUNDATION


Microsoft


ALFRED P. SLOAN FOUNDATION
1934


UNIVERSITY of WASHINGTON
eScience Institute


Urban@UW


CASCADIA URBAN
ANALYTICS COOPERATIVE