Defining Type of Cyclist Based on Travel Activity Patterns

2018 National Household Travel Survey Workshop

Jim Sullivan, Sarah Howerter
UVM Transportation Research Center

Defining Type of Cyclist Based on Travel Activity Patterns

- Motivation
- Method
- Results
- Conclusions

Motivation

- Typical cyclists classifications have been based on driver classifications – like commuters / non-commuters
- Efforts to promote cycling commonly focus on personal and household characteristics of cyclists
- Since cycling is by nature more social than driving, cyclists often exhibit travel characteristics that transcend personal and household characteristics
- Travel behavior, by time of day and purpose, can be utilized to classify cyclists using methods common to marketing and website analytics (also known as *behavioral cohort analysis*)

Method

- Adapted from:
  - Cui, Yu, Qing He, and Alireza Khani, 2018. “Travel Behavior Classification: An Approach with Social Network and Deep Learning.” Transportation Research Record 2018, pp. 1–13. Published online June 18, 2018.
  - Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, 2008. "Fast unfolding of communities in large networks." Journal of statistical mechanics: theory and experiment (2008): P10008.

Method

- Person-trip data for persons who cycled on their travel day in the 2017 NHTS is used to classify cyclists
- For each person, a binary matrix is constructed with trip purposes and times of day on the axes, and a 1 in each cell if the activity was taking place.
- The similarity of every pair of these binary matrices are calculated (Jaccard similarity coefficient) and presumed linkages are drawn between respondents based on cells in the binary matrices that match, and the strength of the linkage between any two respondents is proportional to the number of matches in their binary matrices
- From Cui et al, 2018:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}$$

Method

- Scores below a threshold for Jaccard similarity coefficient are set to 0, effectively "disconnecting" those respondents
- Induced networks are built and network clusters are inferred (Blondel et al, 2008) using the concept of *Modularity*
- The characteristics of the new clusters can be examined as a new behavior-based classification

Results

- 3,606 persons in the 2017 NHTS made 8,034 bike trips on their travel day

	HBW	HBSHOP	HBSOCREC	HBO	NHB	Total
Weekday	1,128	793	1,729	1,474	1,255	6,379
Weekend	88	310	597	247	413	1,655
Total	1,216	1,103	2,326	1,721	1,668	8,034

Results

- Binary matrices:

	T00:00	T00:01	T00:02	T00:03	T00:04	T00:05	T00:06	T00:07	T00:08	T00:09	...	T23:50	T23:51	T23:52	T23:53	T23:54	T23:55	T23:56	T23:5
HBW	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
HBSHOP	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
HBSOCREC	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
HBO	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
NHB	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	

5 rows × 1440 columns

Results

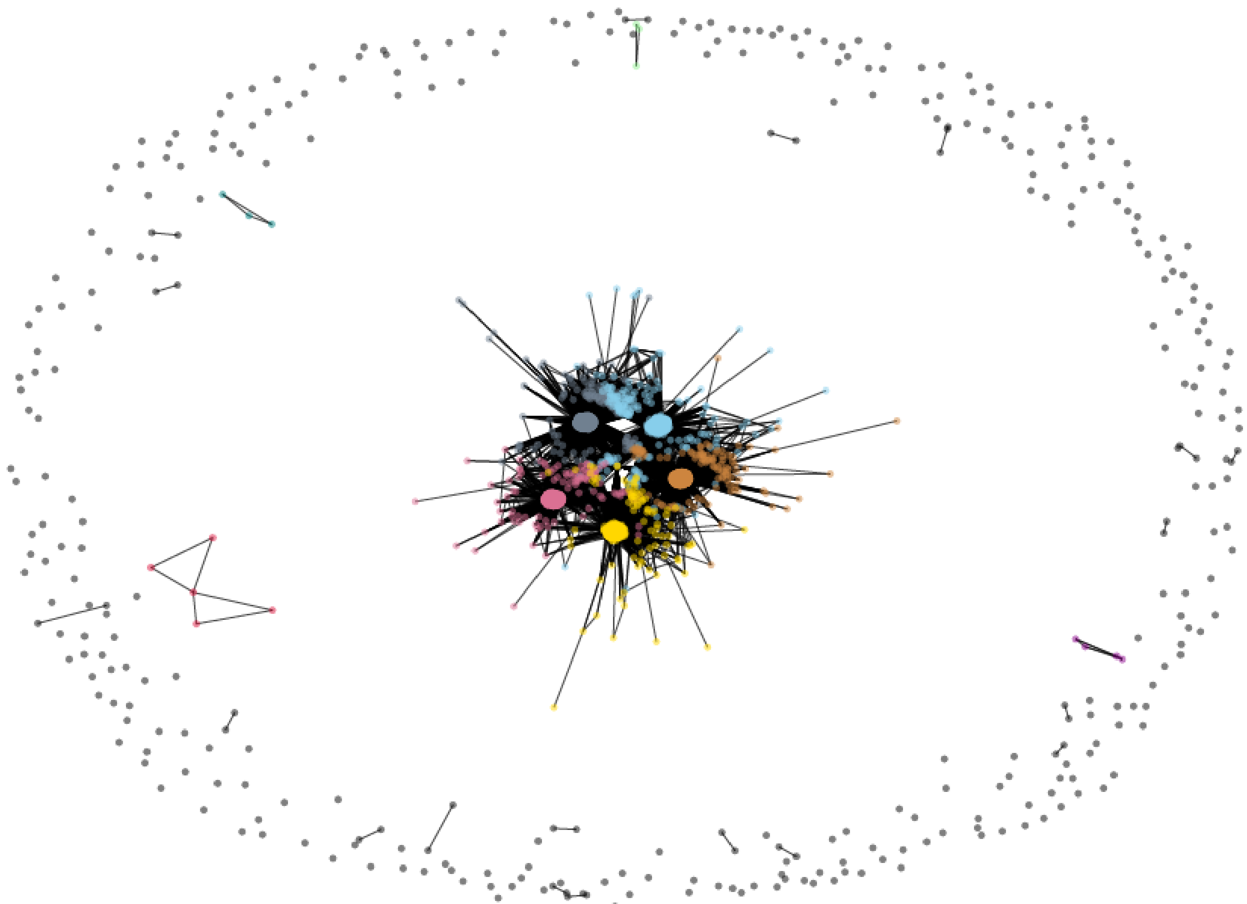
	3050509401	3034911101	3037343002	3028581902	3037403601	3010291901	3031627701	3045228301	4007580002	300
3050509401	NaN	0.6	0.866667	0.8	0.2	0.8	0.6	0.6	0.8	
3034911101	0.6	NaN	0.6	0.6	0.414815	0.6	0.8	0.6	0.6	
3037343002	0.866667	0.6	NaN	0.8	0.2	0.8	0.6	0.6	0.8	
3028581902	0.8	0.6	0.8	NaN	0.2	0.8	0.6	0.6	0.8	
3037403601	0.2	0.414815	0.2	0.2	NaN	0.2	0.4	0.2	0.2	
3010291901	0.8	0.6	0.8	0.8	0.2	NaN	0.6	0.6	0.8	
3031627701	0.6	0.8	0.6	0.6	0.4	0.6	NaN	0.6	0.6	
3045228301	0.6	0.6	0.6	0.6	0.2	0.6	0.6	NaN	0.6	
4007580002	0.8	0.6	0.8	0.8	0.2	0.8	0.6	0.6	NaN	
3005620903	0.6	0.6	0.6	0.6	0.4	0.6	0.6	0.6	0.6	
4066621002	0.6	0.4	0.6	0.6	0.2	0.627273	0.4	0.4	0.6	
3026638201	0.4	0.6	0.4	0.4	0.4	0.4	0.621374	0.4	0.4	
3034992501	0.6	0.6	0.6	0.6	0.4	0.6	0.6	0.6	0.6	
3051508702	0.6	0.6	0.6	0.6	0.2	0.6	0.6	0.8	0.6	
3009155303	0.6	0.6	0.6	0.6	0.4	0.6	0.6	0.6	0.6	
3012470802	0.8	0.6	0.8	0.811538	0.2	0.816667	0.6	0.6	0.8	
3022332503	0.6	0.6	0.6	0.6	0.4	0.6	0.6	0.6	0.6	
3005586302	0.6	0.6	0.6	0.6	0.4	0.6	0.6	0.6	0.6	(
3033793603	0.6	0.6	0.6	0.6	0.4	0.6	0.6	0.6	0.6	

Results

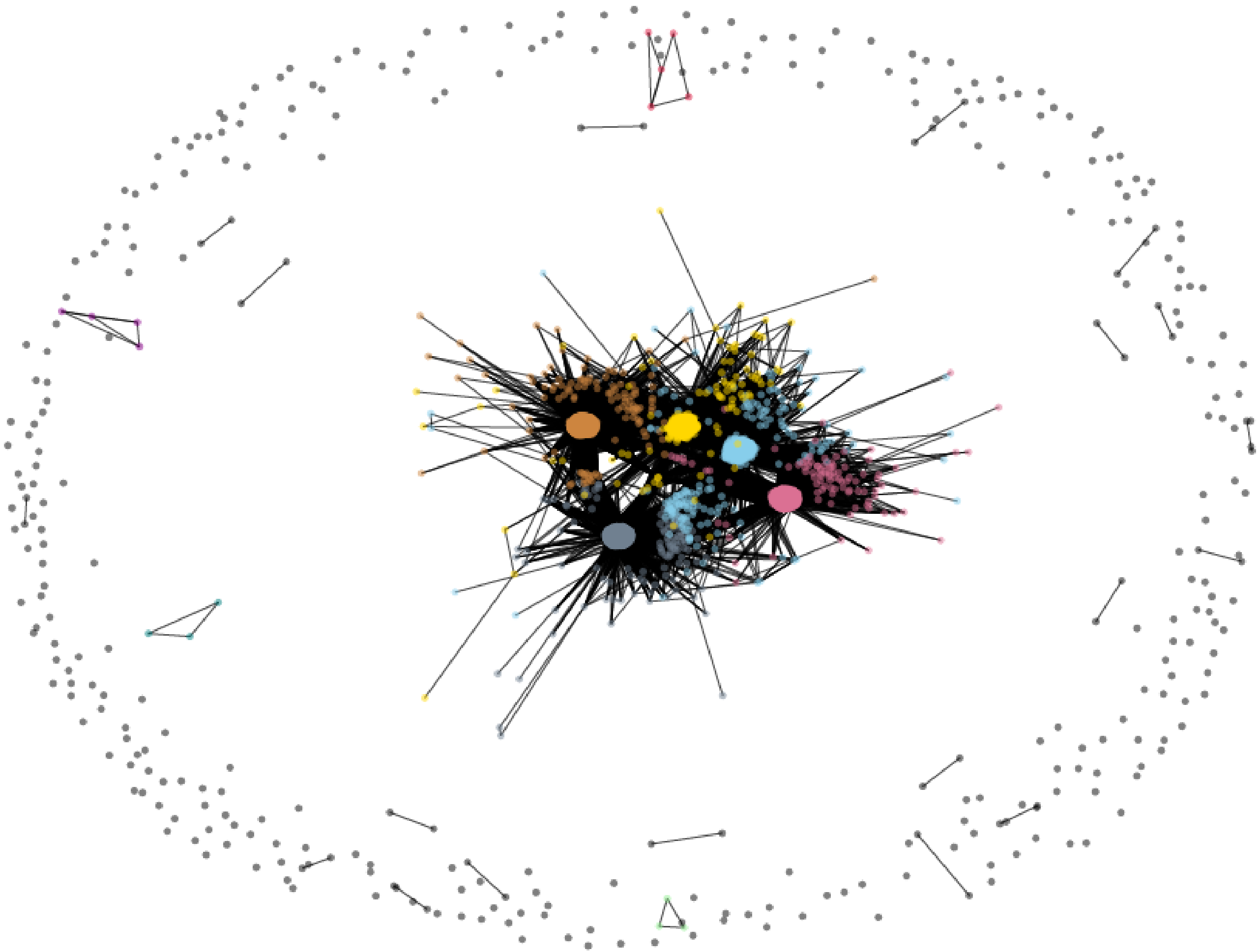
With a similarity threshold of 0.65:

	3050509401	3034911101	3037343002	3028581902	3037403601	3010291901	3031627701	3045228301	4007580002	300
3050509401	0.000000	0.0000	0.866667	0.800000	0.0	0.800000	0.00	0.0	0.800000	0
3034911101	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.80	0.0	0.000000	0
3037343002	0.866667	0.0000	0.000000	0.800000	0.0	0.800000	0.00	0.0	0.800000	0
3028581902	0.800000	0.0000	0.800000	0.000000	0.0	0.800000	0.00	0.0	0.800000	0
3037403601	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3010291901	0.800000	0.0000	0.800000	0.800000	0.0	0.000000	0.00	0.0	0.800000	0
3031627701	0.000000	0.8000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3045228301	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
4007580002	0.800000	0.0000	0.800000	0.800000	0.0	0.800000	0.00	0.0	0.000000	0
3005620903	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
4066621002	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3026638201	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3034992501	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3051508702	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.8	0.000000	0
3009155303	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3012470802	0.800000	0.0000	0.800000	0.811538	0.0	0.816667	0.00	0.0	0.800000	0
3022332503	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3005586302	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0
3033793603	0.000000	0.0000	0.000000	0.000000	0.0	0.000000	0.00	0.0	0.000000	0

Results



Results



Results

Cluster groupings in the induced network for a variety of Similarity Thresholds

Cluster	N by Similarity Threshold		
	0.65	0.70	0.75
1	1,010	966	935
2	443	451	337
3	528	505	454
4	914	895	848
5	351	322	282
6	3	2	2
7	2	3	2
8	2	3	3
9	2	2	3
….	1-5	1-5	1-5

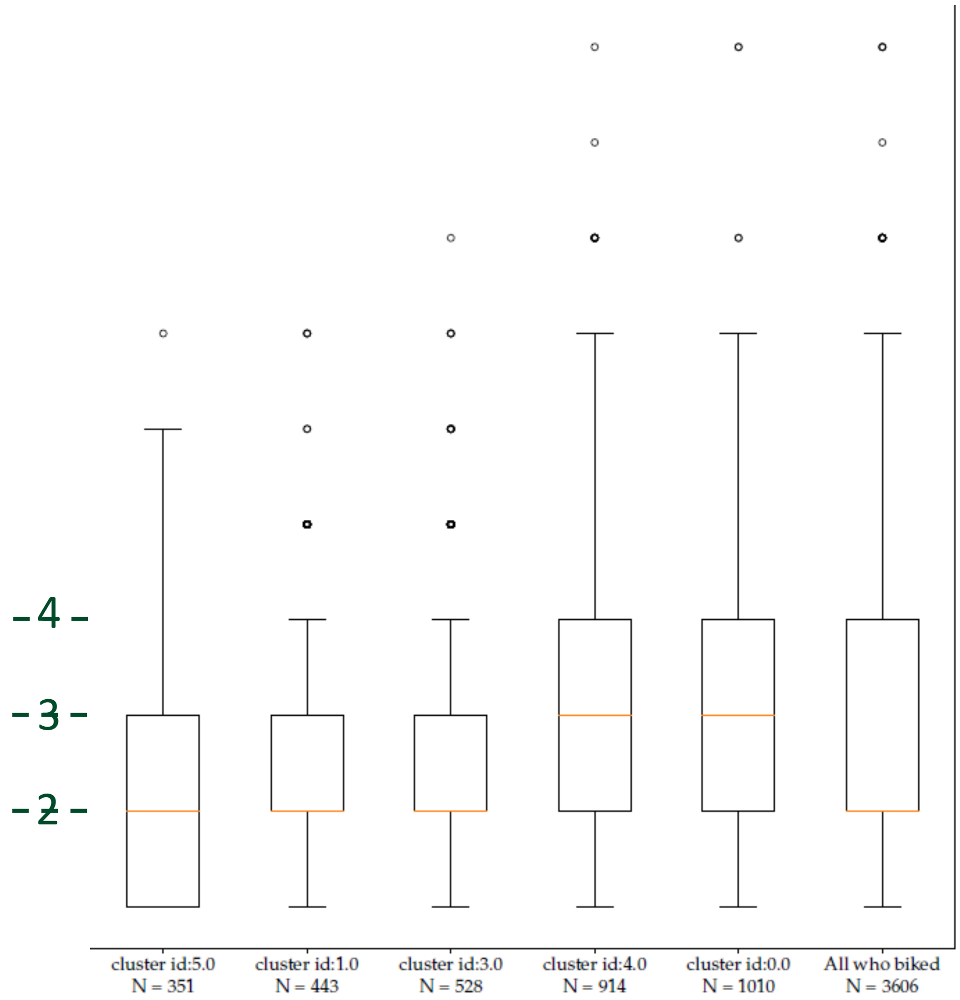
Results

- The demographic characteristics of these clusters can compared:

Results

Household Size

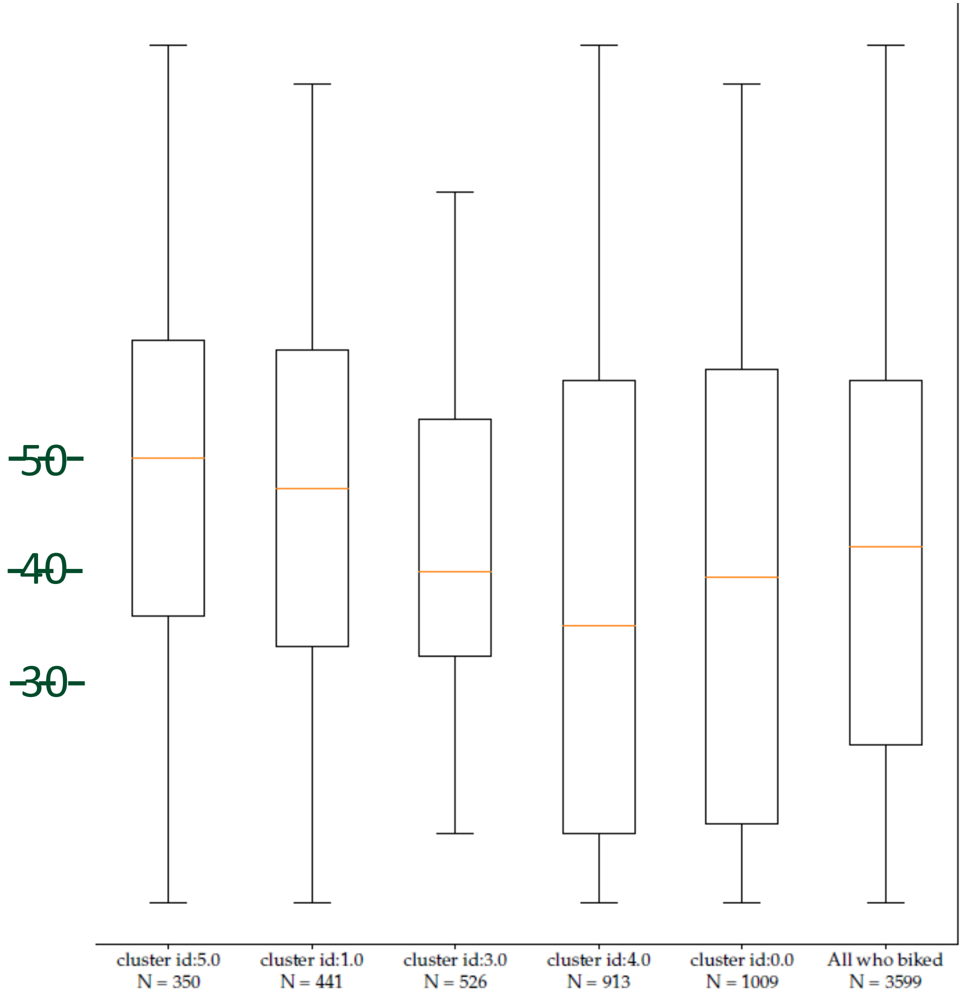
- Orange line is the median
- Box is the 25th and 75th percentiles
- Bars are the range (outliers are dots)



Results

Age

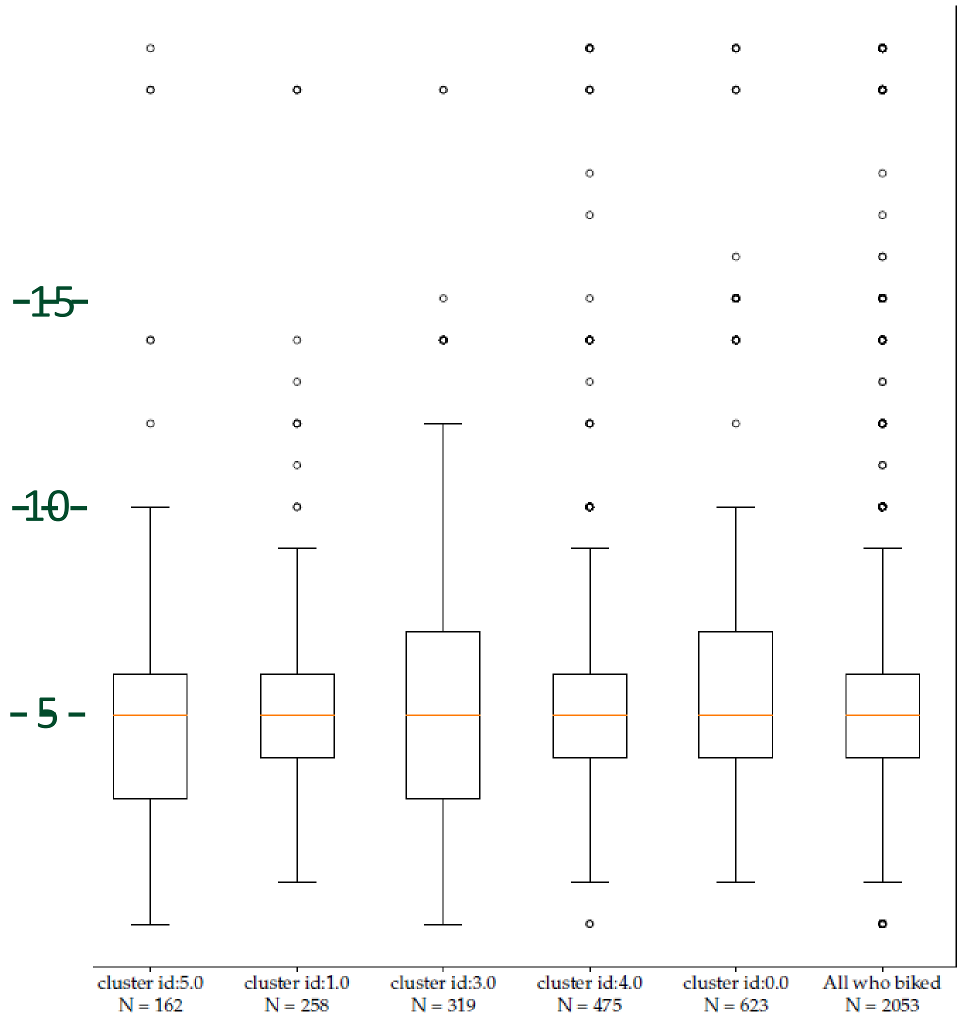
- Orange line is the median
- Box is the 25th and 75th percentiles
- Bars are the range (outliers are dots)



Results

Vigorous Activity in the Past Week

- Orange line is the median
- Box is the 25th and 75th percentiles
- Bars are the range (outliers are dots)



Results

Average Age by Home Ownership for Primary Behavioral Clusters

Cluster	Own	Rent	All
1	39.1	32.1	**37.5**
2	48.7	36.8	**45.5**
3	45.3	35.0	**40.7**
4	39.0	27.6	**36.1**
5	48.5	44.9	**47.1**
All	**42.1**	**34.0**	**39.7**

Conclusions

- This method allows us to characterize cyclists according to their behavior, without regard to who they are, or what their household looks like
- We can then retroactively describe the clusters according to the social associations that are present

Conclusions

- In marketing, these types of characterizations, developed based on purchasing behavior and web use, result in "Super Clusters" that help businesses market to consumers:
  - Upper Crust
  - High Fidelity
  - Net Worth & Networks
  - Picket Fences
  - Maintaining a Balance
  - Ways & Means
  - Golden Years
  - Debt Builders
  - Hardscrabbles

Questions?

- What will our clusters be called?
  1.
  2.
  3.
  4.
  5.