A New Way to Measure Clustering Where It Matters: Evidence of Retail Agglomeration at a Very Small Scale

Leah Brooks ▪ Rachel Meltzer ▪ May 2026

Are Consumers the Future of Cities?

- Cities grew as hubs for production
- Production left cities and cities grew as knowledge hubs
- Role of cities as knowledge hub may be shrinking?

Are Consumers the Future of Cities?

- Cities grew as hubs for production
- Production left cities and cities grew as knowledge hubs
- Role of cities as knowledge hub may be shrinking?
- Some argue that cities are now consumption hubs

Measuring Cities as Consumption Hubs

- Need a way to measure spatial organization of consumer-facing activity
- Need to know: do consumer-facing establishments cluster more than other types of establishments?
- And, how does consumer-facing clustering vary by establishment type, location or the demographics of nearby residents?

Answer with a New Measure and Big Data

Measure

- Count establishments within x meters of a focal establishment
- An establishment-specific measure of clustering
- Fix the distance, characterize the distribution of clustering rather than a industry-specific summary statistic

Answer with a New Measure and Big Data

Measure

- Count establishments within x meters of a focal establishment
- An establishment-specific measure of clustering
- Fix the distance, characterize the distribution of clustering rather than a industry-specific summary statistic

Data

- Source: DataAxle
- Every establishment in the US
- Today: 2019 only
- Entire dataset: back to late 1990s
- In data center: back to 1982

Agenda for Today

1. What can our new measure explain that existing measures do not?
2. Are consumer-facing establishments more clustered than random? Than other industries?
3. Characterization of consumer-facing clustering at a small scale

Why Do We Need Our Measure?

We Have Tools to Measure Industry Agglomeration

- Ellison-Glaeser Index (1997): Compare industry location to "dartboard" randomness
- Duranton-Overman Index (2005): Exploits pairwise distances

We Have Tools to Measure Industry Agglomeration

- Ellison-Glaeser Index (1997): Compare industry location to "dartboard" randomness
- Duranton-Overman Index (2005): Exploits pairwise distances
- Primarily applied to production-oriented firms
- Both rely on pre-determined geography and create industry-level measures

Why We Need an Establishment-level Cluster Metric: Intuition

What are the underlying economics that lead to small scale clustering?

- **Consumer-facing**: any establishment that relies on in-person contact for purchase of services or goods.

Why We Need an Establishment-level Cluster Metric: Intuition

What are the underlying economics that lead to small scale clustering?

- **Consumer-facing**: any establishment that relies on in-person contact for purchase of services or goods.
- **Establishments** locate to optimize between fixed costs (rent) and distance to consumer base.
- **Consumers** choose establishments to optimize travel and search costs
- **Clustering** reduces travel and search costs for the consumer

Why We Need an Establishment-level Cluster Metric: Intuition

What are the underlying economics that lead to small scale clustering?

- **Consumer-facing**: any establishment that relies on in-person contact for purchase of services or goods.
- **Establishments** locate to optimize between fixed costs (rent) and distance to consumer base.
- **Consumers** choose establishments to optimize travel and search costs
- **Clustering** reduces travel and search costs for the consumer
- Limits to time spent "consuming" and frequency of consumption drive size and spatial regularity of clusters

Duranton-Overman Index: A Canonical Measure for Clustering

We observe some estabs. in space

Estab.	lat.	long.
A	3	3
B	2	1
C	1	2

Duranton-Overman Index: A Canonical Measure for Clustering

We observe some estabs. in space

Estab.	lat.	long.
A	3	3
B	2	1
C	1	2

Find pairwise distances between all estabs.

Estab. 1	Estab. 2	dist.
A	B	2.2
A	C	2.2
B	C	1.4

Duranton-Overman Index: A Canonical Measure for Clustering

We observe some estabs. in space

Estab.	lat.	long.
A	3	3
B	2	1
C	1	2

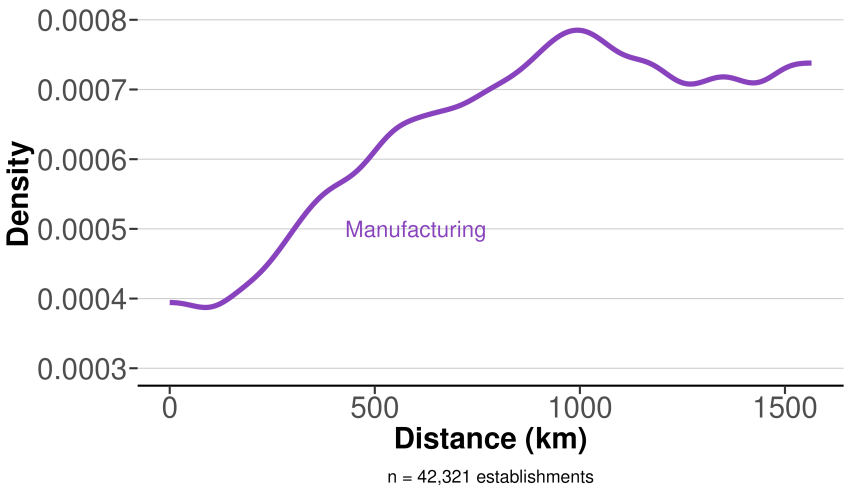
Find pairwise distances between all estabs.

Estab. 1	Estab. 2	dist.
A	B	2.2
A	C	2.2
B	C	1.4

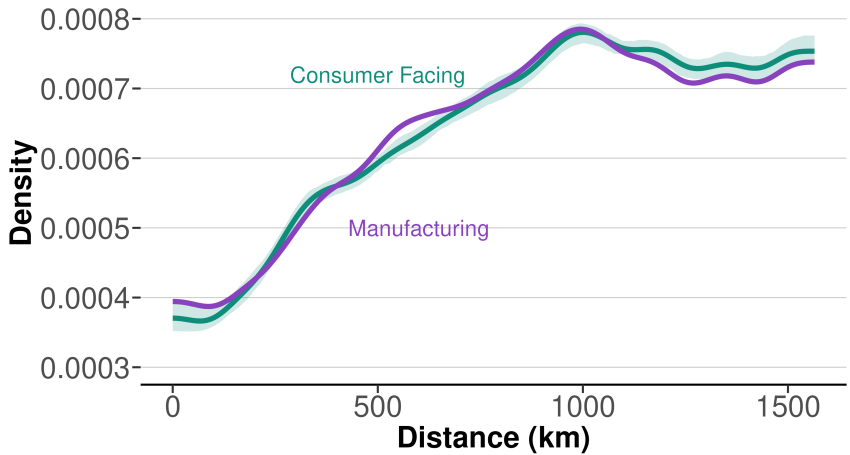
$\rightarrow$ Plot distribution of pairwise distances.

Mostly Long Pairwise Distances btwn Manufacturing Estabs. in the US

Based on ten percent random sample of all manufacturing establishments

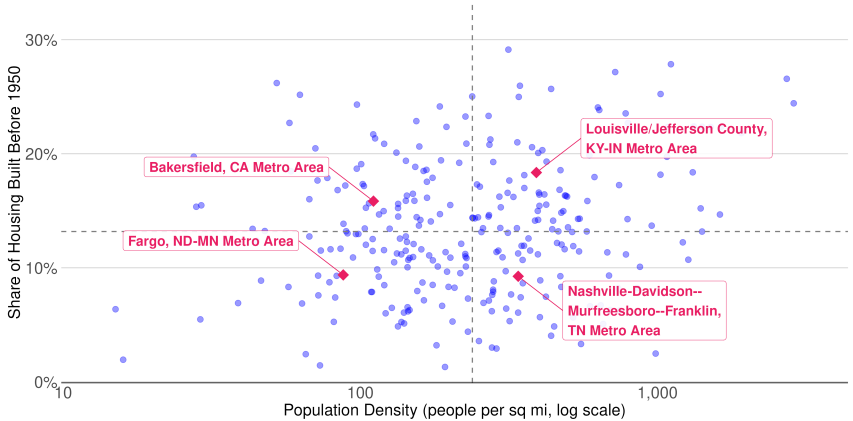


Distance Distribution for Consumer-facing Tracks Manufacturing



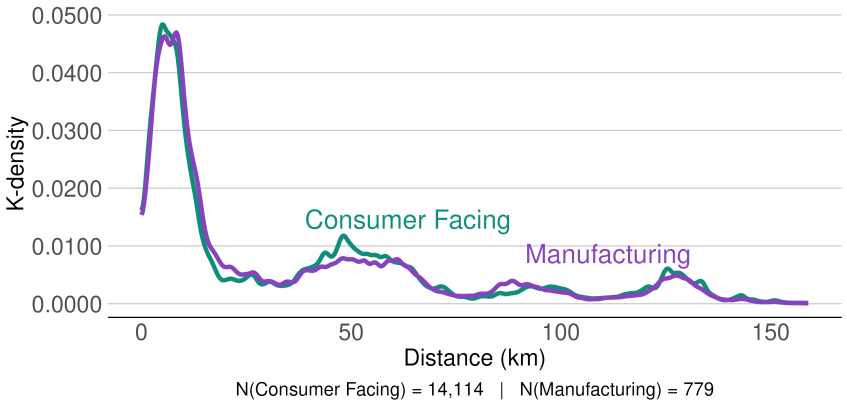
CF: 100 replications of 5,000 establishments | MFG: 42,321 establishments

Drill Down to MSAs

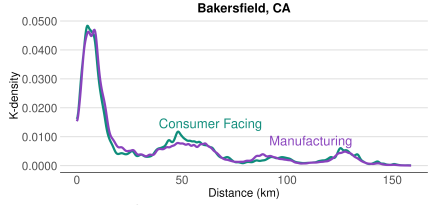


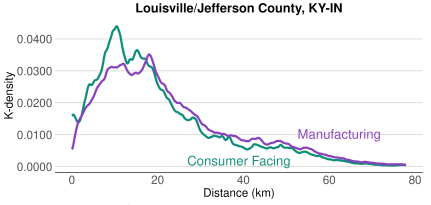
Still Little Difference in Pairwise Distance Distribution

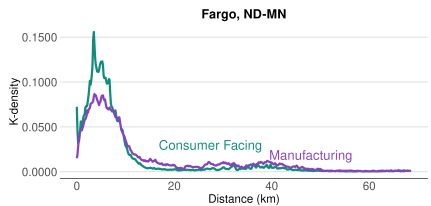
Bakersfield, CA

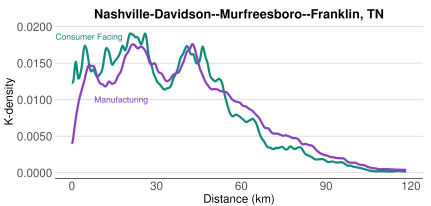


Across MSAs, Consumer-facing and Manufacturing Distance Distributions Mostly Track



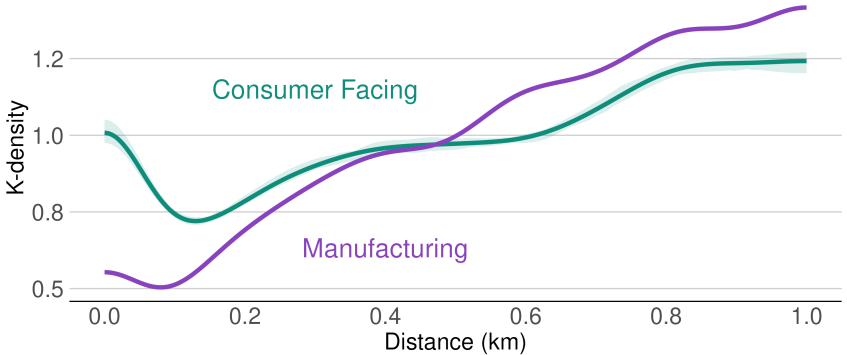




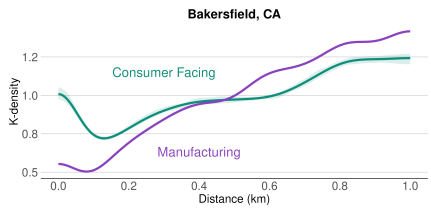


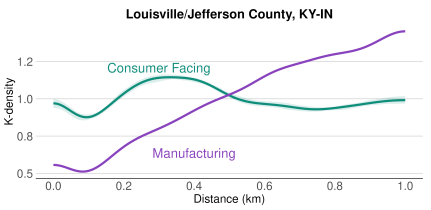
Curves Diverge at Pairwise Distances Less than 1 km

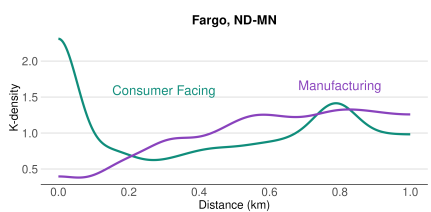
Bakersfield, CA

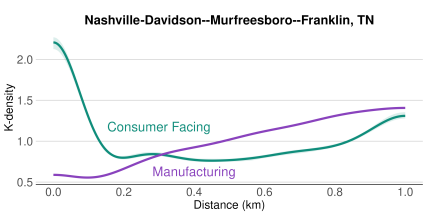


Across MSAS, Curves Diverge at Pairwise Distances Less than 1 km

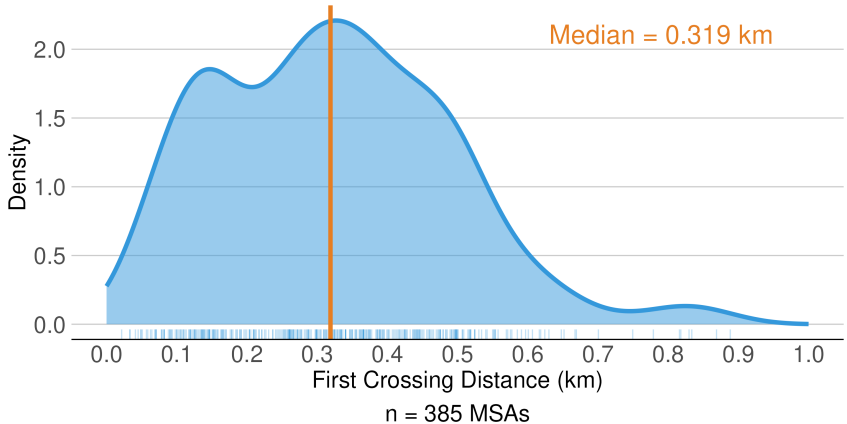








Median Crossing is 300m Across All MSAs



What Can We Add to Existing Clustering Indices?

- Existing indices measure industry-wide tendency of establishments to cluster

What Can We Add to Existing Clustering Indices?

- Existing indices measure industry-wide tendency of establishments to cluster
- We want to know
  - clustering at level of individual establishment

What Can We Add to Existing Clustering Indices?

- Existing indices measure industry-wide tendency of establishments to cluster
- We want to know
  - clustering at level of individual establishment
  - full distribution of establishment clustering at small scales and for any one locality
  - distribution of in establishment clustering by sub-industry
  - how clustering varies across space
  - how clustering changes over time (not today)

What Does Our Measure Show?

Introducing and Applying the Establishment-level Measure

Introducing the Measure:

1. Absolute clustering
2. Test if different from random
3. Normalized clustering
4. Compare consumer-facing and manufacturing clustering

Introducing and Applying the Establishment-level Measure

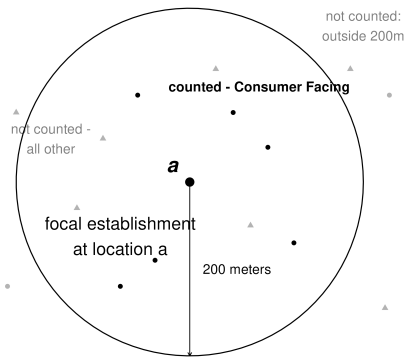
Introducing the Measure:

1. Absolute clustering
2. Test if different from random
3. Normalized clustering
4. Compare consumer-facing and manufacturing clustering

Applying the Measure:

1. Does consumer-facing clustering change across the distribution?
2. Are larger clusters specialized or diverse?
3. Does clustering change over space?

Calculating An Establishment-level Cluster Metric



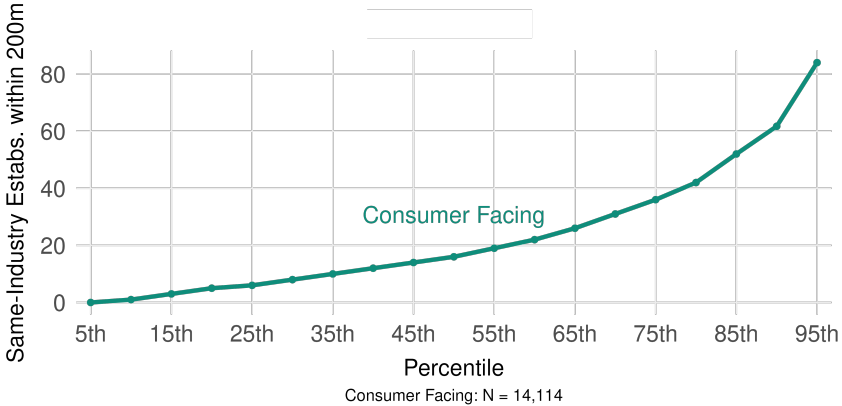
- Establishment i, industry m
- d_{ij}: distance between establishment i and any other establishment $j \neq i$
- we fix a distance
- maximum distance for counting r

For establishment i in industry m:

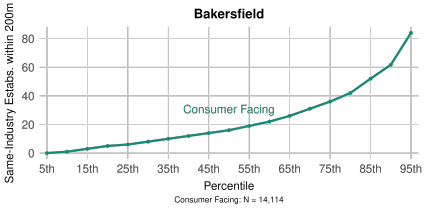
$$C_{i,m} = \sum_{j \neq i} I(d_{ij} < r \text{ and } j \text{ in } m)$$

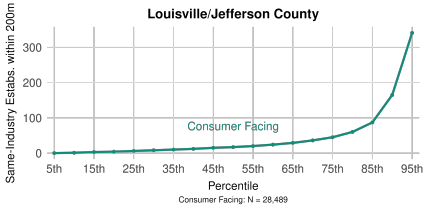
There is a within-MSA Distribution of Consumer-facing Clustering

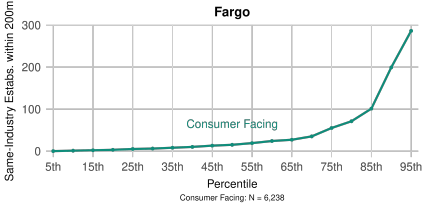
Bakersfield, CA

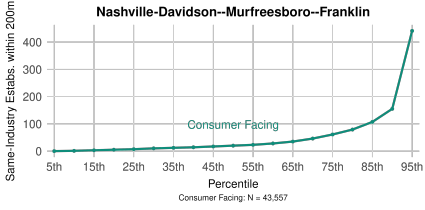


Other MSAs Show Sharp Rise in Clustering









Do Consumer-facing Establishments Cluster More than Random?

Goal: Compare observed clustering to clustering at random commercial locations

- Let N be total number of consumer facing establishments
- Draw N establishments without replacement from all commercial establishment locations
- Find clustering $C_{i,m}$
- Find quantiles of this distribution $C_{i,m}^{(q)}$
- Repeat 100 times
- Report 5th and 95th percentile of counterfactual distribution at each quantile q

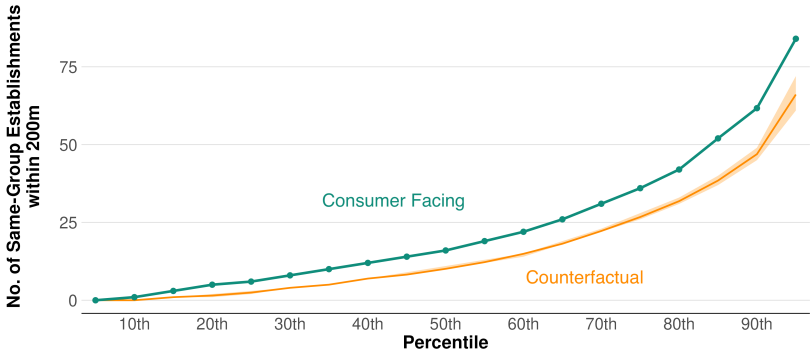
Do Consumer-facing Establishments Cluster More than Random?

Goal: Compare observed clustering to clustering at random commercial locations

- Let N be total number of consumer facing establishments
- Draw N establishments without replacement from all commercial establishment locations
- Find clustering $C_{i,m}$
- Find quantiles of this distribution $C_{i,m}^{(q)}$
- Repeat 100 times
- Report 5th and 95th percentile of counterfactual distribution at each quantile q
- **If observed value exceeds upper bound of counterfactual distribution, establishments are more than randomly clustered**

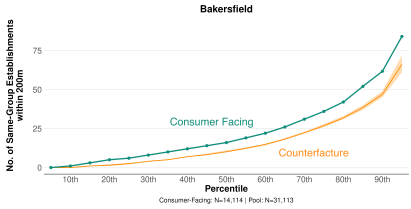
Consumer-facing Establishments Clustering Exceeds Random Locations

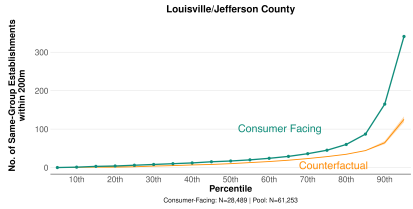
Bakersfield, CA

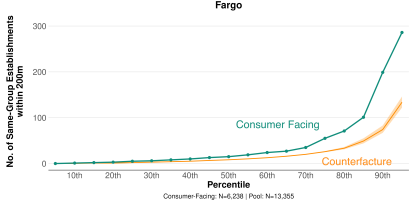


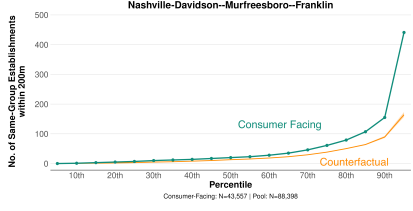
Consumer-Facing: N=14,114 | Pool: N=31,113

Consumer-facing Establishments Clustering Exceeds Random Locations

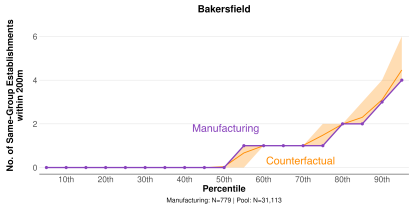


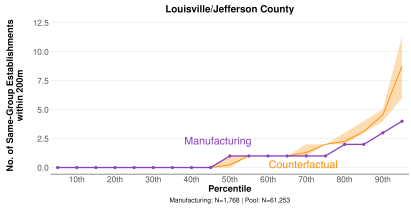


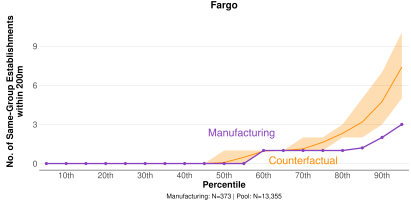


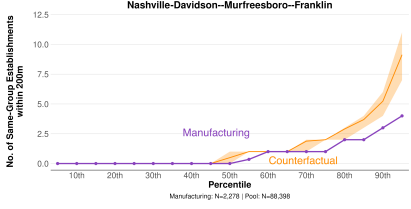


Manufacturing Establishments are Less Clustered than Random



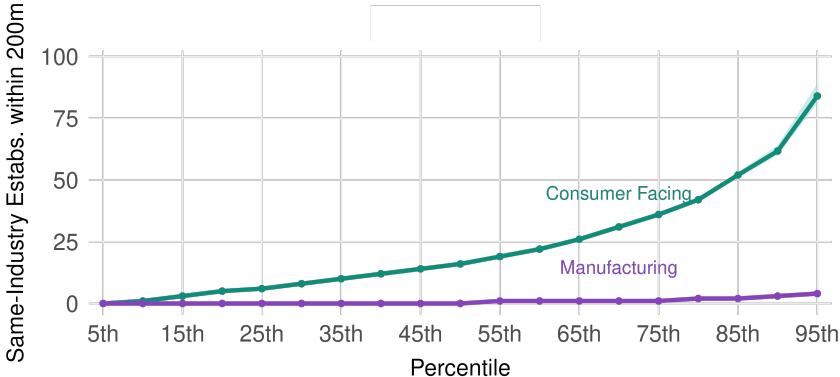






Comparing Consumer-facing and Manufacturing

Bakersfield, CA



Consumer-facing: N = 14,114 | Manufacturing: N = 779

Accounting for Differently Sized Industries

- But measure $C_{i,m}$ depends on number of establishments in industry m!
- To compare across industries, we must correct for this

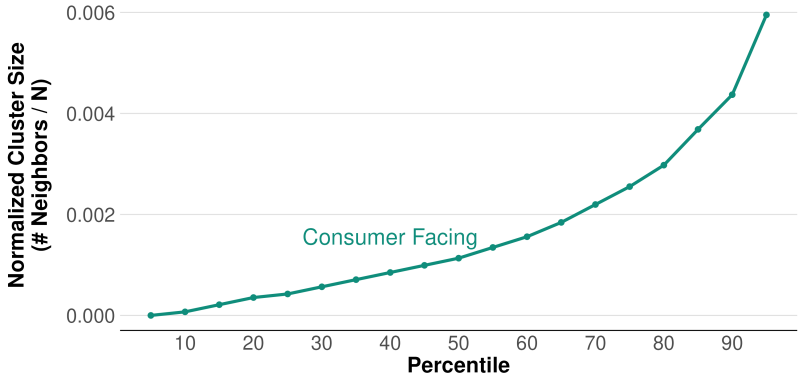
Accounting for Differently Sized Industries

- But measure $C_{i,m}$ depends on number of establishments in industry m!
- To compare across industries, we must correct for this
- $C_{N,m}$: total number of establishments in industry m
- Create normalized clustering measure for establishment i in industry m:

$$c_{i,m} = \frac{C_{i,m}}{C_{N,m}}$$

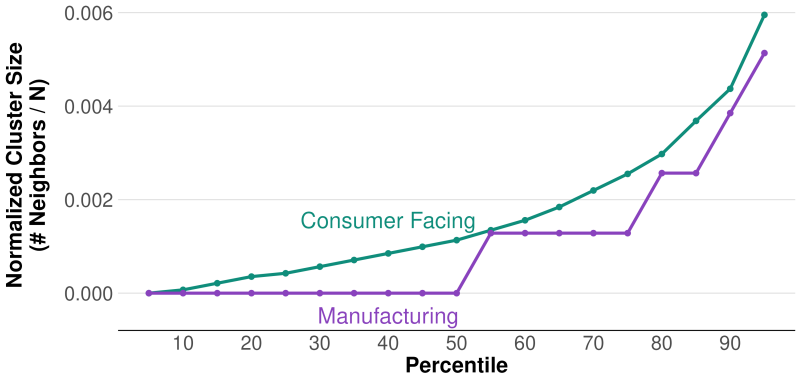
Normalized Consumer-facing Clustering Looks Like Absolute Clustering

Bakersfield, CA

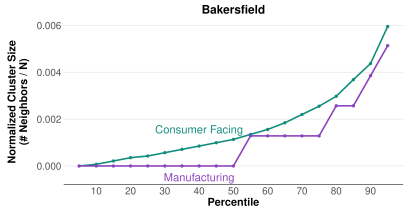


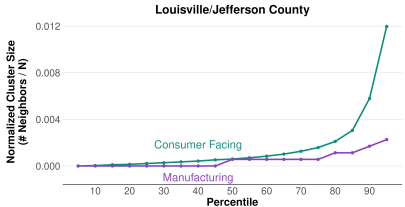
Consumer-facing Establishments Cluster More than Manufacturing

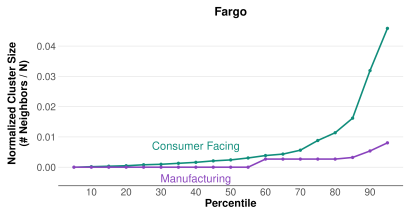
Bakersfield, CA

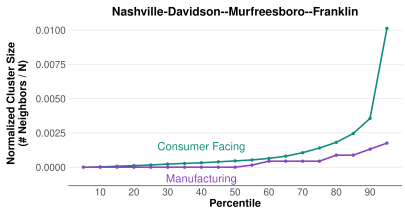


Consumer-facing Establishments Cluster More than Manufacturing

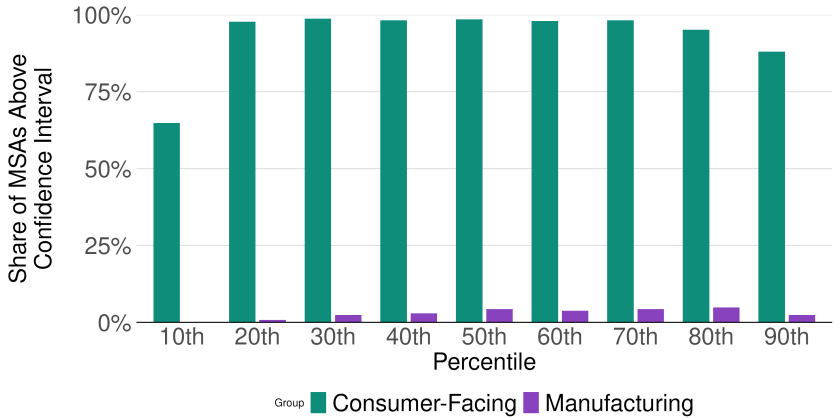








Consumer-facing Establishments More Clustered Across All MSAs, Across Entire Distribution



Does Normalized Clustering Differ by Percentile?

For example: **is the median CF establishment more clustered relative to other industries compared to the 90th pctl CF establishment?**

q is quantile

M is total number of industries

Ratio of normalized CF clustering to average clustering in all other 2-digit industries:

$$R_m^{(q)} = \frac{c_m^{(q)}}{\sum_{m \neq \text{ consumer facing}} c_m^{(q)} / (M-1)}$$

Does Normalized Clustering Differ by Percentile?

For example: **is the median CF establishment more clustered relative to other industries compared to the 90th pctl CF establishment?**

q is quantile

M is total number of industries

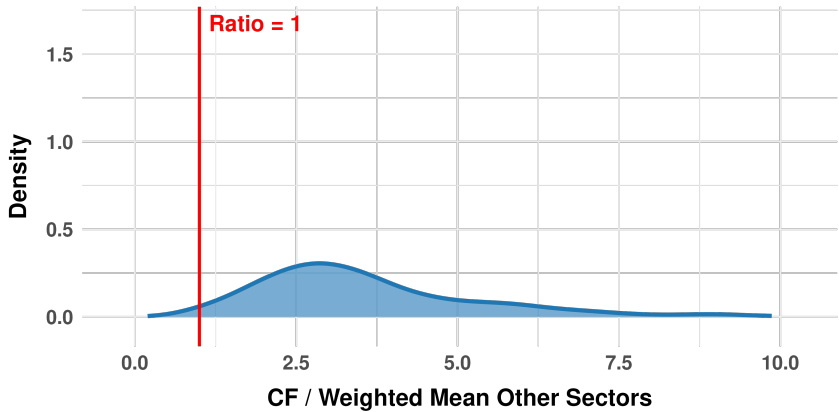
Ratio of normalized CF clustering to average clustering in all other 2-digit industries:

$$R_m^{(q)} = \frac{c_m^{(q)}}{\sum_{m \neq \text{ consumer facing}} c_m^{(q)}/(M-1)}$$

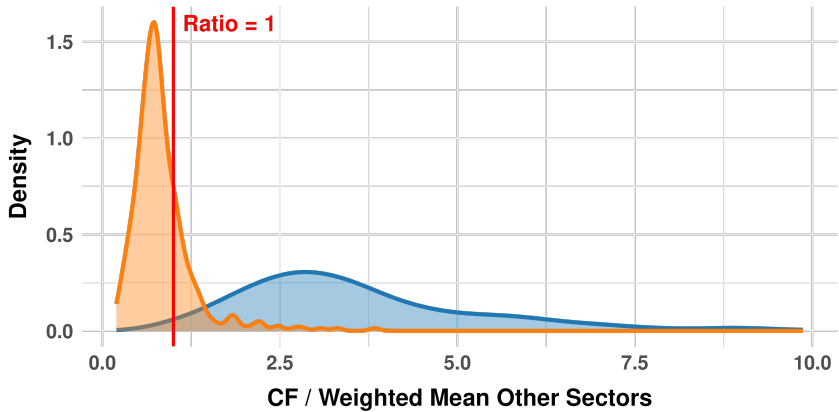
Ratio $> 1 \rightarrow$ median consumer facing establishment more clustered than average industry

Ratio $< 1 \rightarrow$ median consumer facing establishment less clustered than average industry

Median Consumer-facing Establishment More Clustered than Average Industry



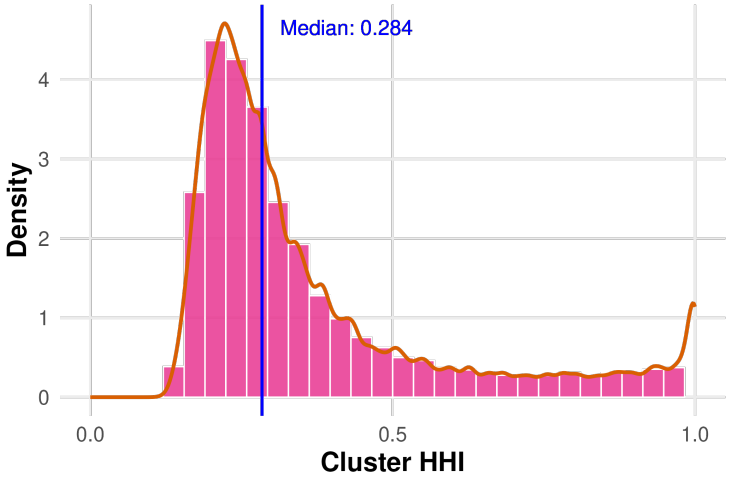
90th Percentile Consumer-facing Establishments Less Clustered than Average Industry



Are the Biggest Establishment-Clusters Specialized or Diverse?

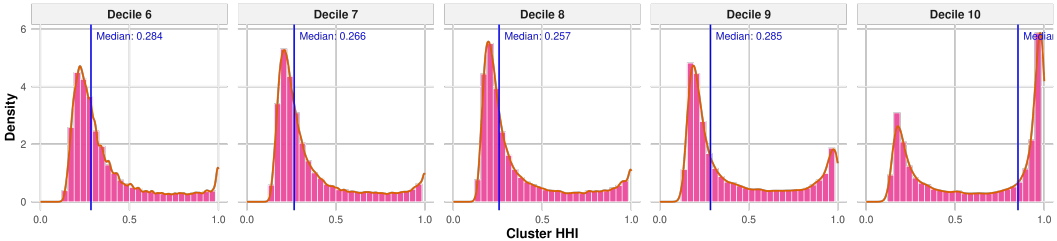
- Measure within cluster industry concentration for neighbors around establishment i
- Use HHI, with shares for two digit consumer facing categories
- K total 2-digit subindustries, denote each as k
- Let S_k denote establishment i's share of neighbors in industry k
- $HHI_i = \sum_{k=1}^{K} S_k^2$
- Values closer to 1 = more concentrated (specialized)

Neighbors of Median Clustered Establishments are Mostly Diverse



N: Decile 6 (n=608,036)

Only the Very Largest Clusters Show Specialization



N: Decile 6 (n=608,048); Decile 7 (n=604,349); Decile 8 (n=608,036); Decile 9 (n=609,776); Decile 10 (n=607,871)

Does Clustering Relate to Neighborhood Attributes?

Adding tract information

- Identify 2010 tracts and neighbors within 10 km
- Find distance-decay weighted attributes
- Join each establishment to attributes of nearby tracts

Does Clustering Relate to Neighborhood Attributes?

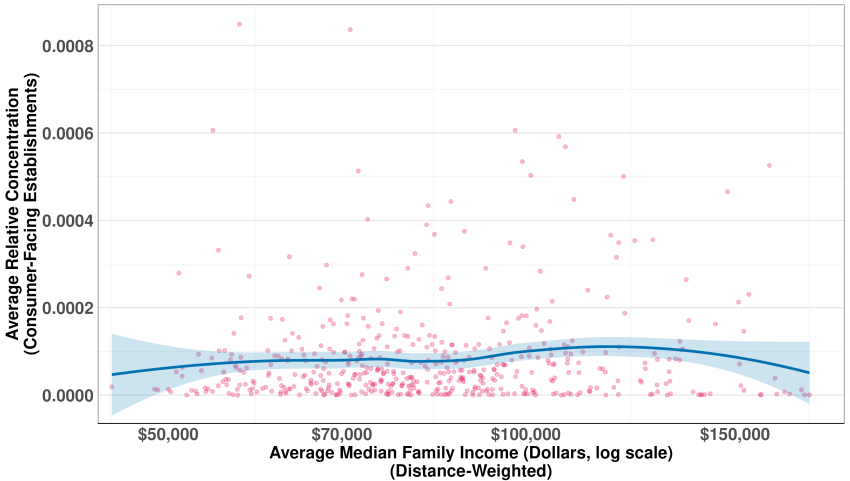
Adding tract information

- Identify 2010 tracts and neighbors within 10 km
- Find distance-decay weighted attributes
- Join each establishment to attributes of nearby tracts

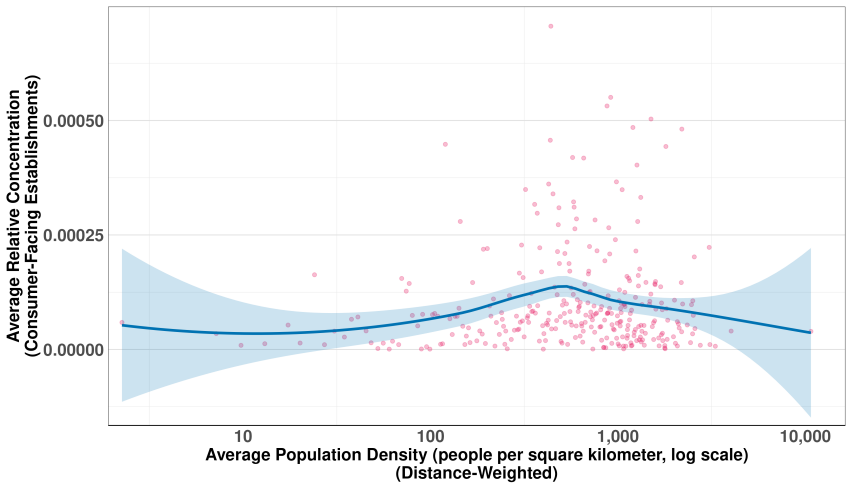
Show relationship w/ binned scatter plot

- Make 300 - 500 bins of the attribute
- In each bin, find average normalized consumer-facing clustering
- Plot relationship between clustering and attribute

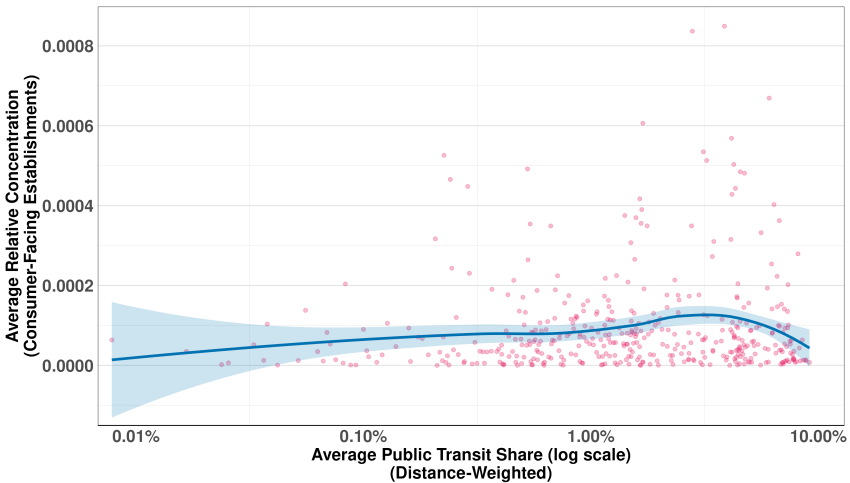
Weak Relationship Between Income and Consumer-facing Clustering



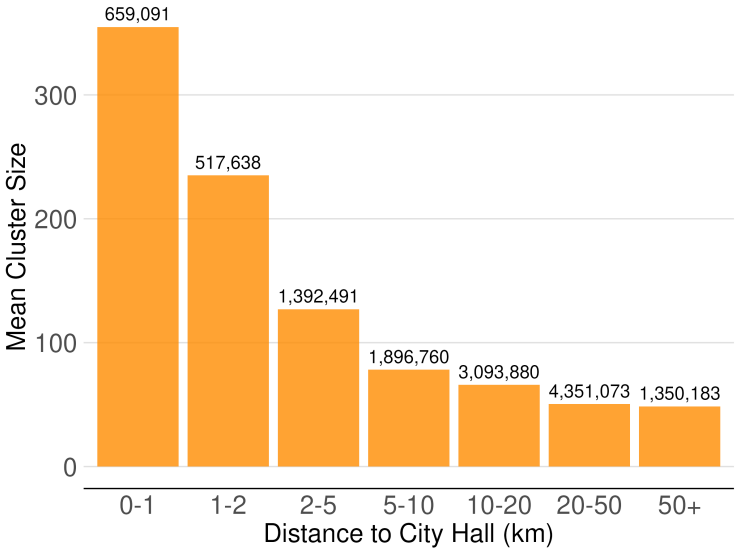
Relationship between Clustering and Population Density is Perplexing



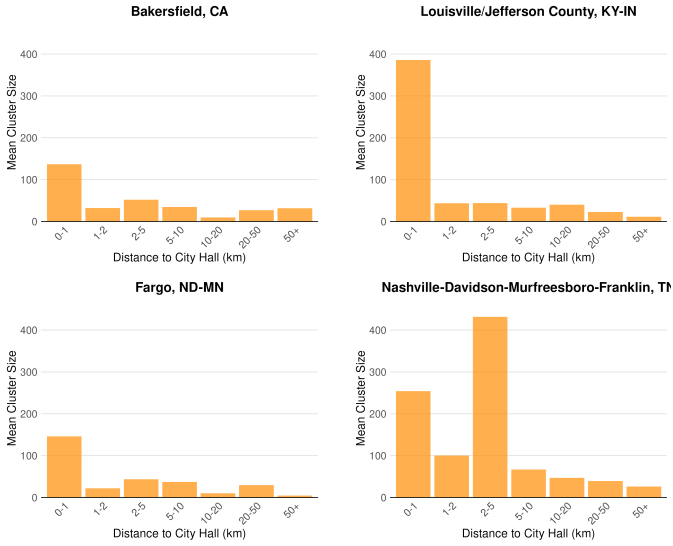
Clustering Increases with Share of Public Transit



Documenting the Allocation of Clusters Across Space, All MSAs



Documenting the Allocation of Clusters Across Space, 4 MSAs



Final Thoughts

- We propose a new agglomeration metric to capture (the distribution of) clustering at small scales
- At short distances, consumer-facing establishments cluster more than random
- The median consumer-facing establishment clusters more than manufacturing and other industries
- Consumer-facing establishments near bigger clusters have neighbors that are more specialized and are closer to city centers
- Unclear how much local population dynamics matter
- Much more to do!

Thank You!