L.A. Food Truck Data Analysis:

Examining Yelp Review Counts and Correlations Across Location Factors

fantastic_foodie_five

Matt Breda, Michelle Risucci, Steven Brown,

Raul Villa, Vijay Rajpurohit

Research Motivation and Summary

- *Research question:* **If you were a food truck operator (or aspiring one), where would the most ideal location be to conduct your business?**
- Using search and location data, is it possible to predict high traffic areas that are more likely to draw in especially review-happy customers?

Hypothesis

- *Hypothesis*: **Higher review counts across food trucks in Los Angeles can be explained at least in part by location-specific factors such as:**
  - Food truck density
  - Population density
  - Median household income
  - Median home price

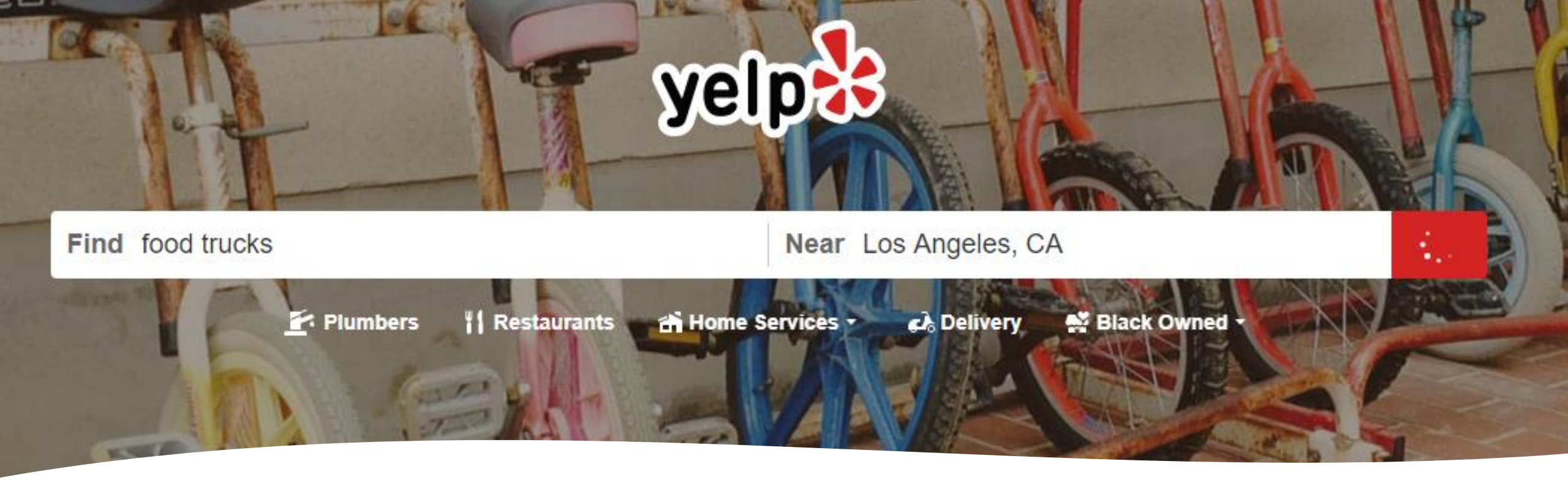
Key Assumptions

- *Key Assumptions*:
  - **Review counts are predictive of business success**, as higher review counts should correspond with higher overall customer volume.
  - **Higher review counts are worth targeting** (over review scores for example), as businesses with higher review counts show up higher in search results and create efficient marketing feedback loops.
  - **Zipcode level demographic information is representative** of the available food truck locations in a given zipcode.
  - **The addresses listed on Yelp are reflective of current business locations.**

Summary Results

Looking at the nearly 600 food trucks in our data set:

- 1) Review counts for food trucks tend to skew very low (generally below 200)
- 2) Demographic variables appear to play a limited role in predicting higher Yelp review counts *for specific food trucks*
- 3) Demographic variables may play a role in explaining *average review counts* by zipcode



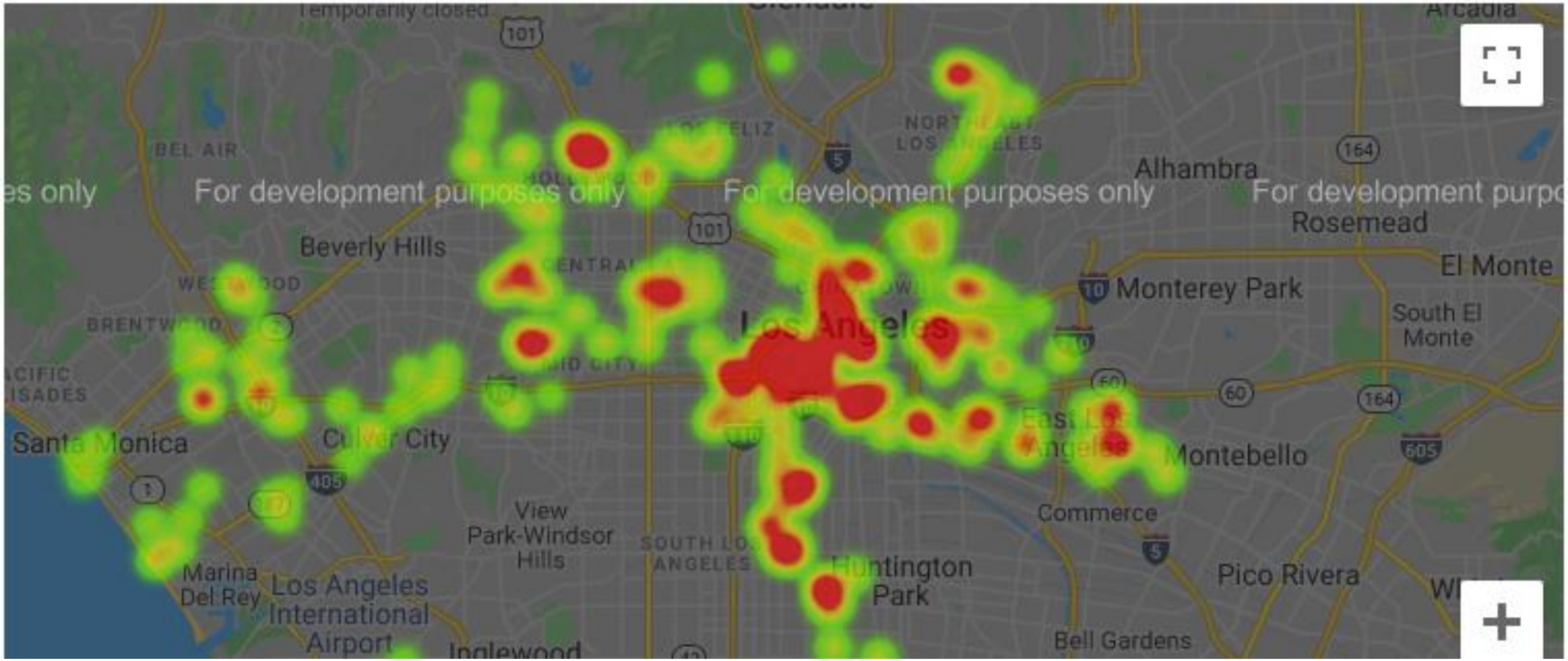
Summary of Data Collection Process

- The Yelp Fusion API was the most helpful in pulling the information needed because we could search by “food truck” as a category.
- The goal was to capture as many unique businesses listed as food truck in order achieve a higher sample.
- By utilizing the PyPi module USZIPCODES, which uses census data, we were able to cross match the zip codes attained from the Yelp API call with Median Home Value, Median Household Income, and Population Density in order to analyze the neighborhood.
- The Google Maps API was helpful for visualizing geographic layout of the food trucks in our data set, using latitude and longitude coordinates from Yelp

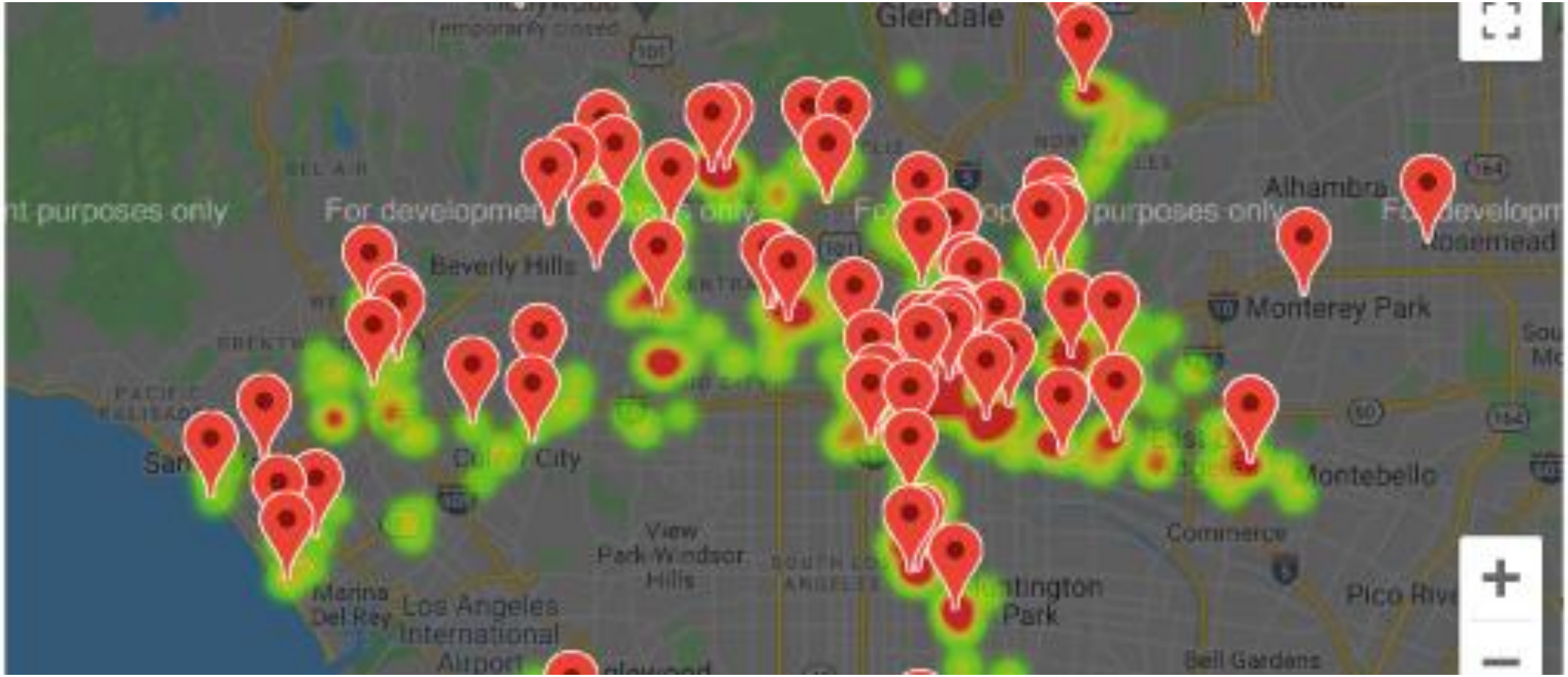
Geographic Layout of Food Trucks in Data Set

- *Research question*: **How are food trucks generally concentrated in L.A. County?**
- Hypothesis: Food trucks tend to be more concentrated in areas that have higher commercial activities, sporting events, or other attractions that drive foot traffic.
- Result: Food trucks are much more concentrated around downtown L.A., and appear to be clustered around many of the city's major freeway intersections, sporting arenas, and shopping areas

Food Truck Density by Zipcode in Los Angeles



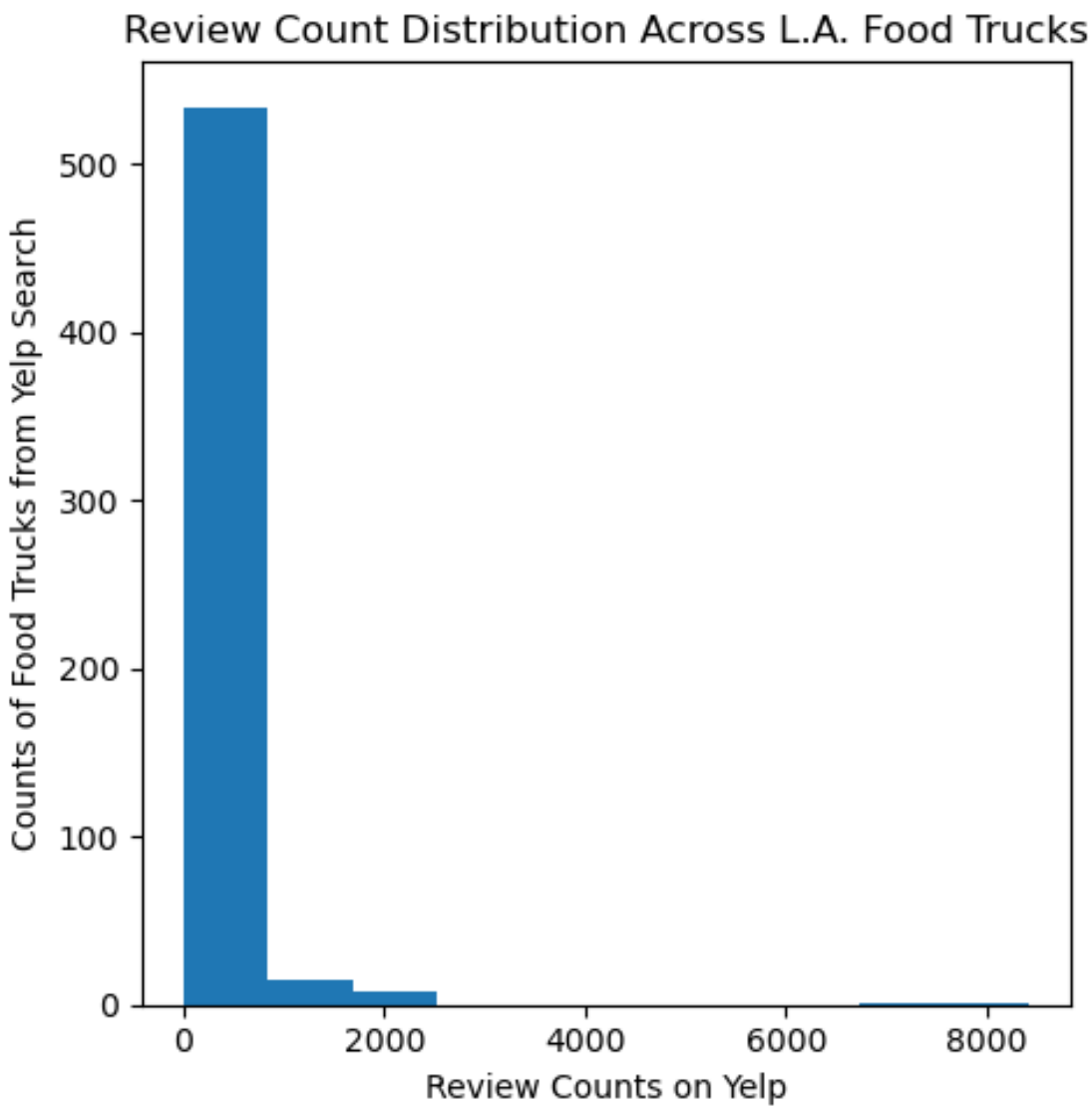
Locations of the Most-Reviewed (Reviews>200) Food Trucks in Los Angeles



Note: The density of the most-reviewed food trucks generally appears to follow overall food truck density.

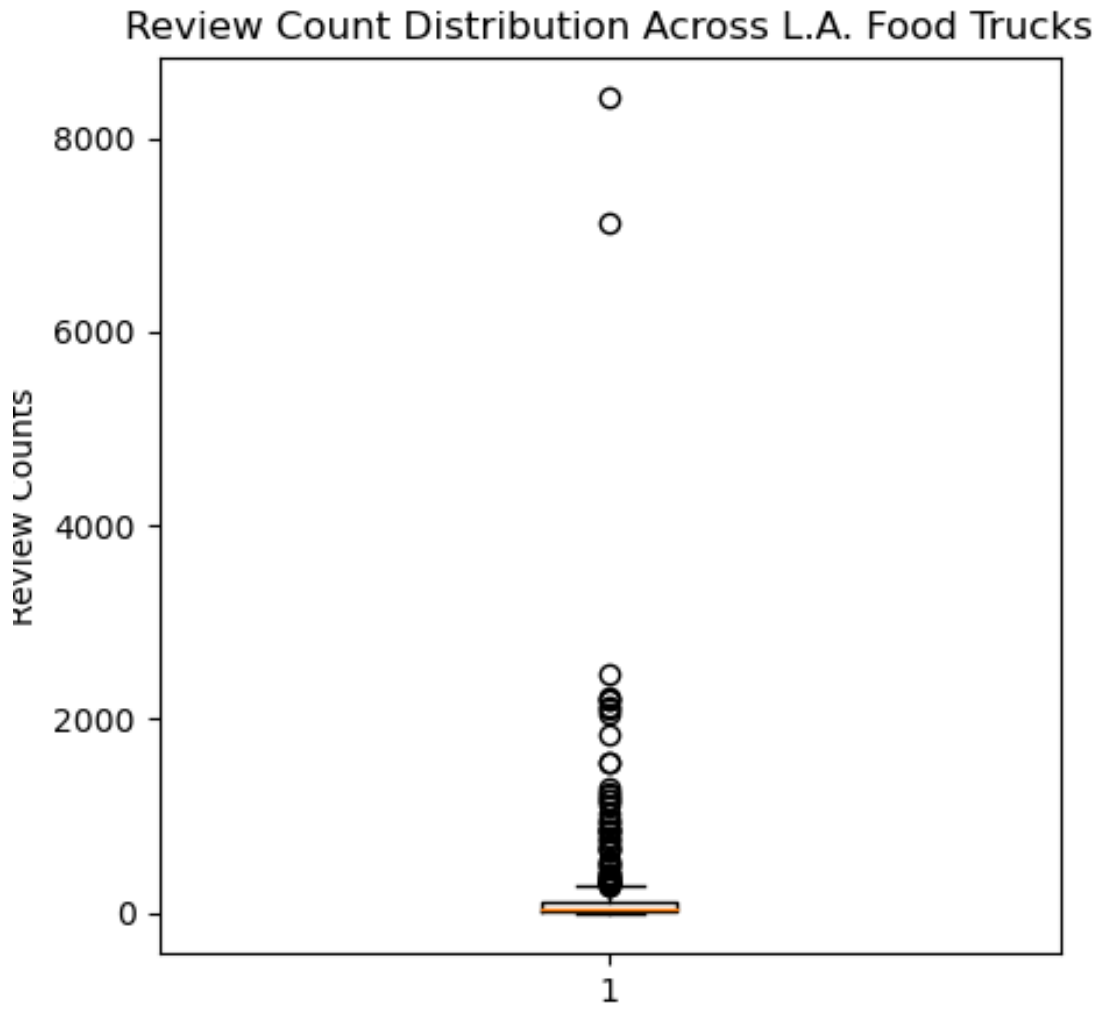
Distribution of Review Counts Across Food Trucks

- *Research question*: **How are review counts distributed among food trucks?**
- Hypothesis: Food truck review counts are likely heavily concentrated among the most instagrammable food trucks with the best locations to drive food traffic.
- Result: Food truck review counts at the are highly concentrated at the high end of the sample; 85% of food trucks in our sample receive 200 reviews or less



Review Count Distribution Across L.A. Food Trucks – Initial Pull from Yelp

n = 559

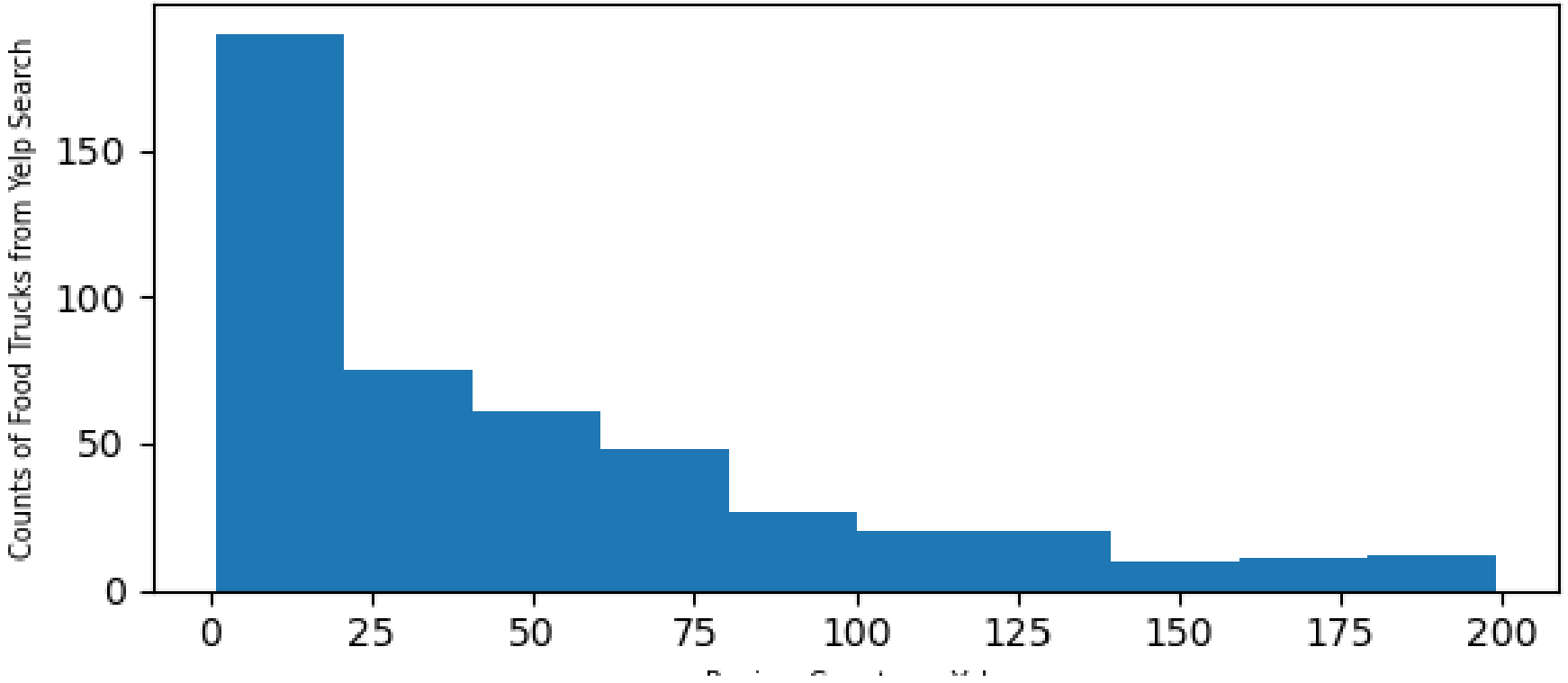


Review Count Distribution Across L.A. Food Trucks – Initial Pull from Yelp

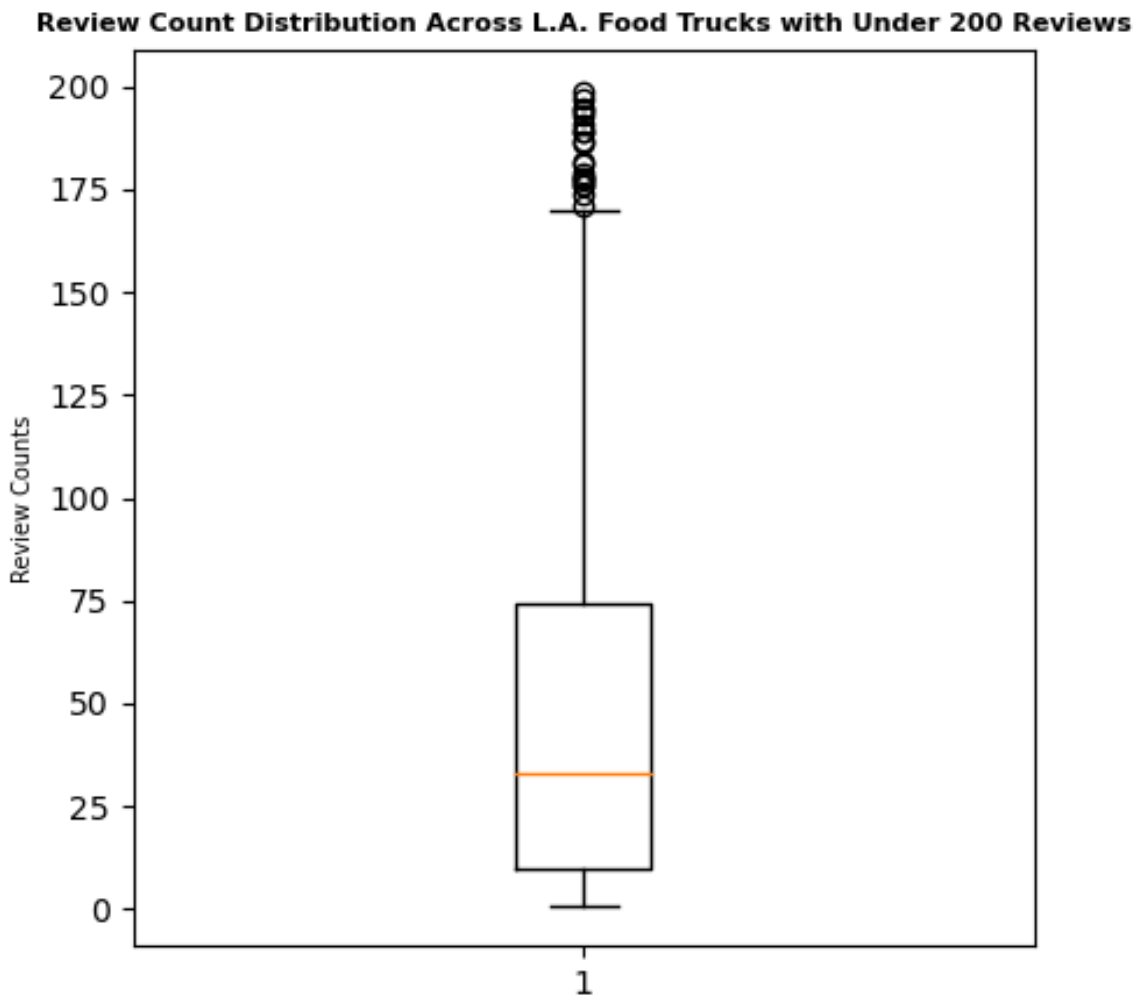
n = 559

Mean: 168
Median: 43
Mode: 2 (count = 17)
Std. Dev: 558
IQR: 12 - 118

Histogram of Review Counts – Filtered for Food Trucks with Under 200 Reviews



n = 474



Box Plot of Review Counts – Filtered for Food Trucks with Under 200 Reviews

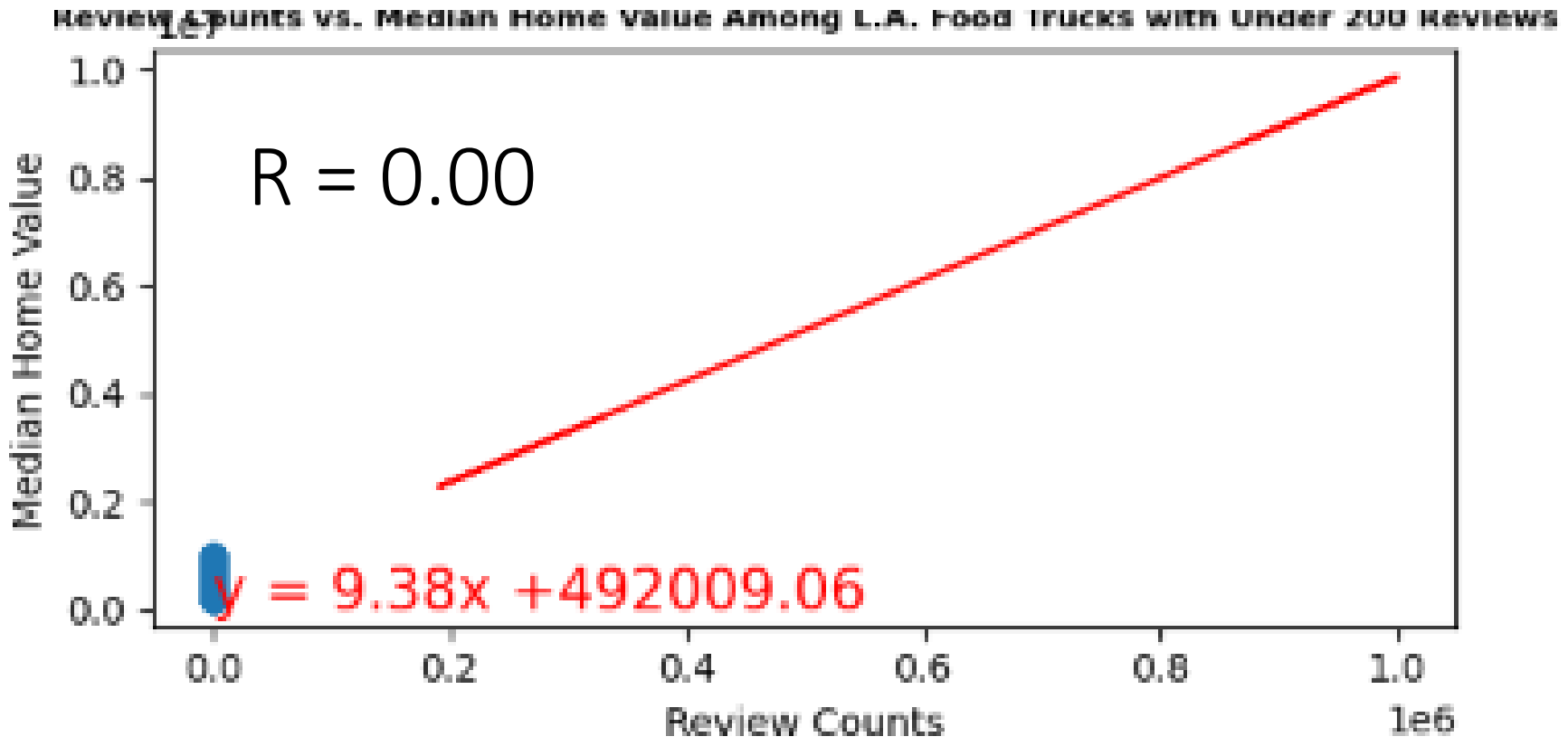
n = 474 (85% of sample)

Mean: 49
Median: 33
Mode: 2 (count = 17)
Std. Dev: 49
IQR: 10 - 74

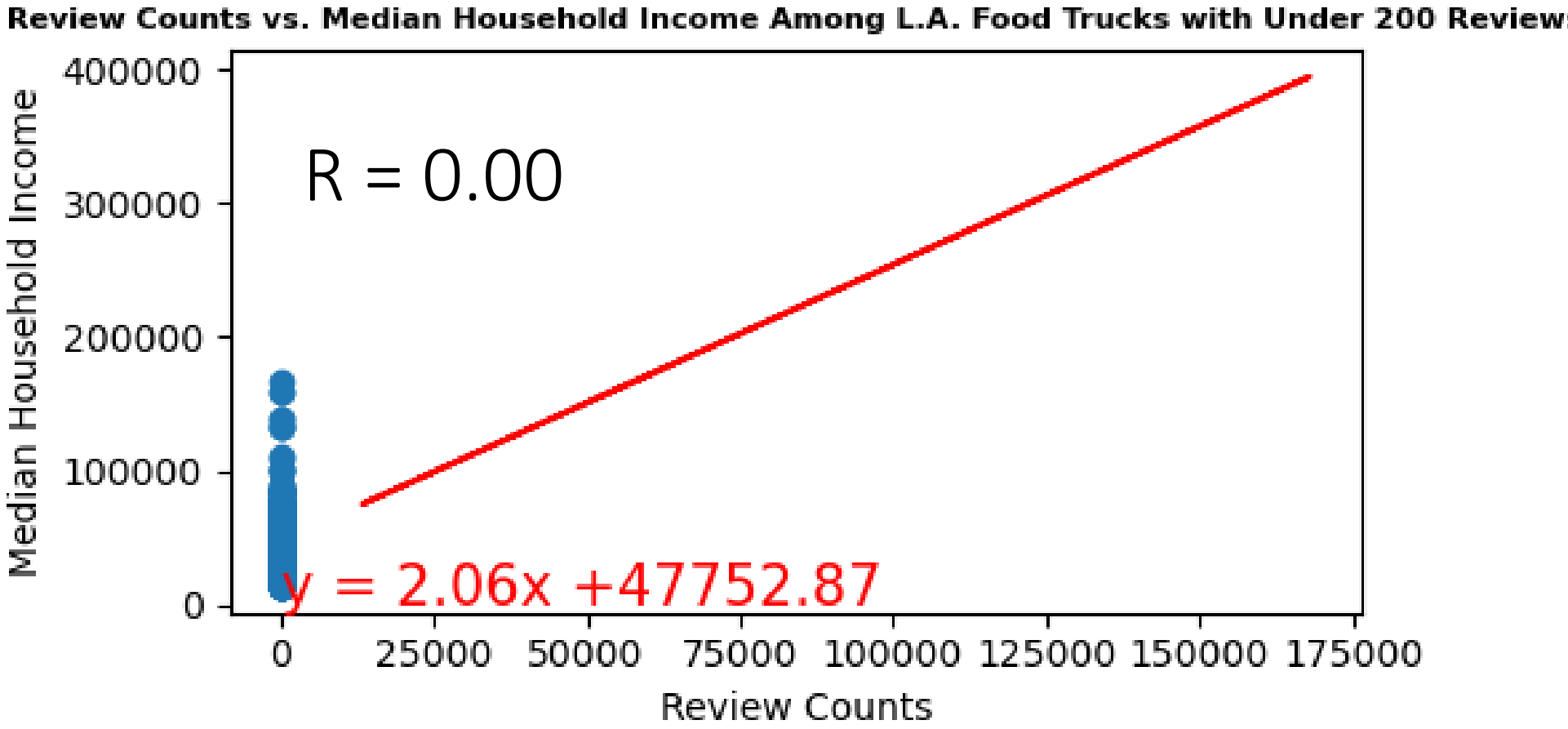
Linear Regression

- *Research question*: **Is there a statistical relationship between a food truck's review count and the demographic features of the zipcode it operates in?**
- Hypothesis: Food truck review counts are likely to be correlated with areas that have high disposable income – e.g. high median home values, high median household incomes, and low population density.
- Result: We found virtually zero statistical relationship between individual food truck review counts and median home values, median household incomes, and population density.

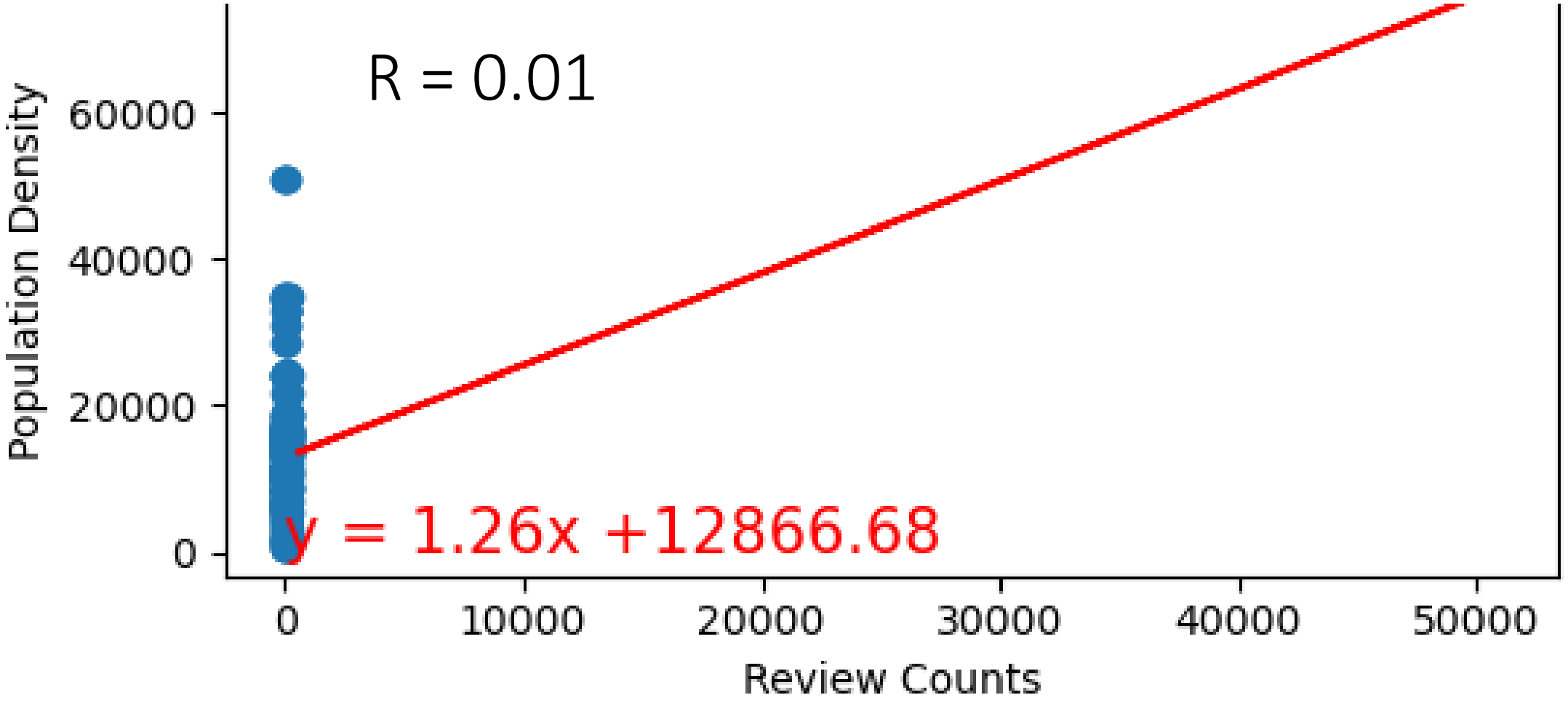
Among L.A. Food Trucks with Under 200 Reviews - Linear Regression: Review Counts vs. Median Home Value



Among L.A. Food Trucks with Under 200 Reviews - Linear Regression: Review Counts vs. Median Household Income

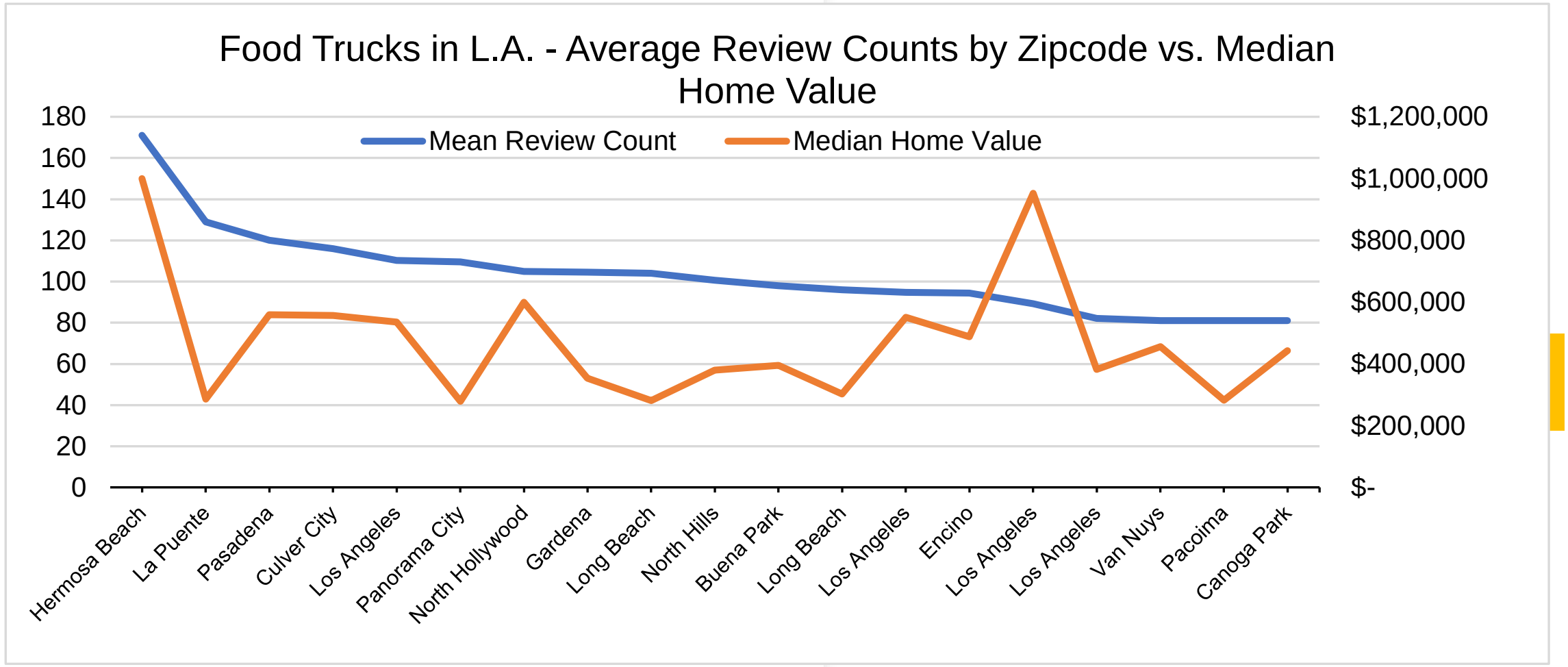


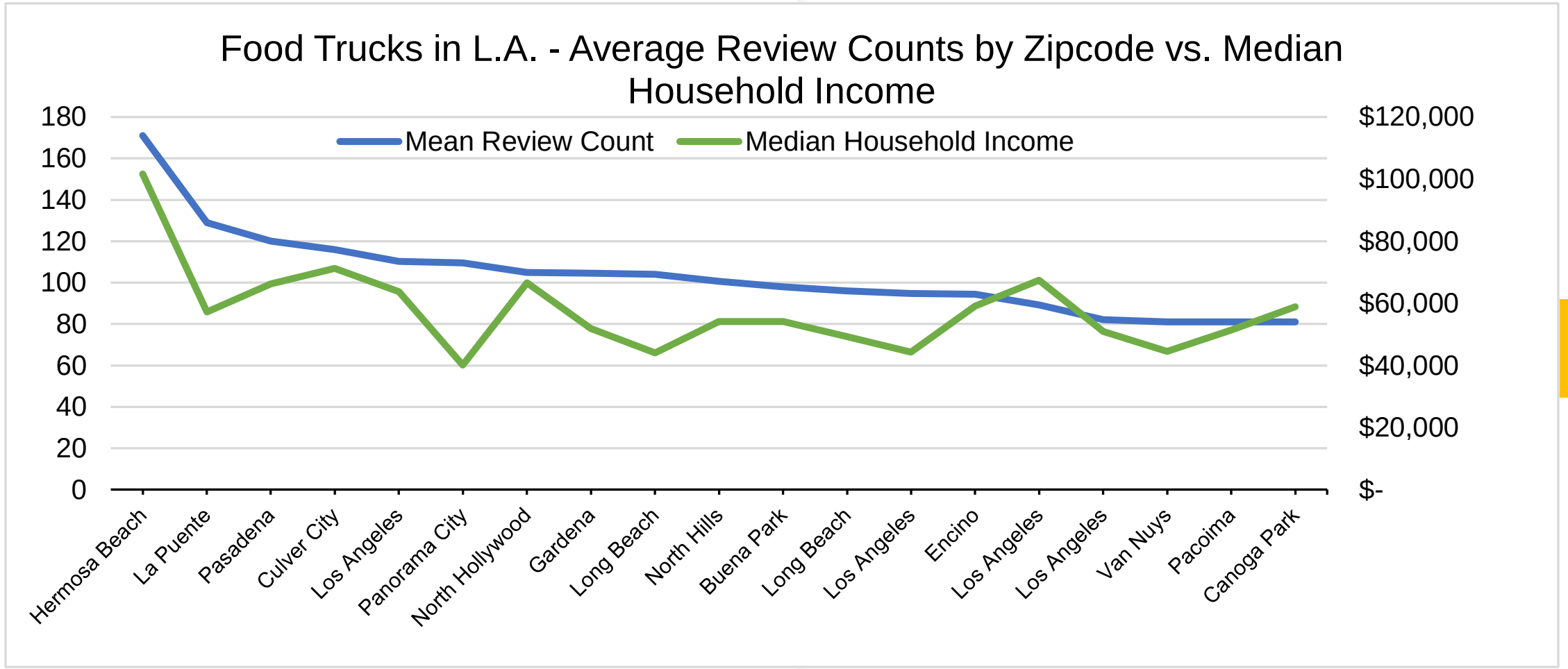
Among L.A. Food Trucks with Under 200 Reviews - Linear Regression: Review Counts vs. Population Density

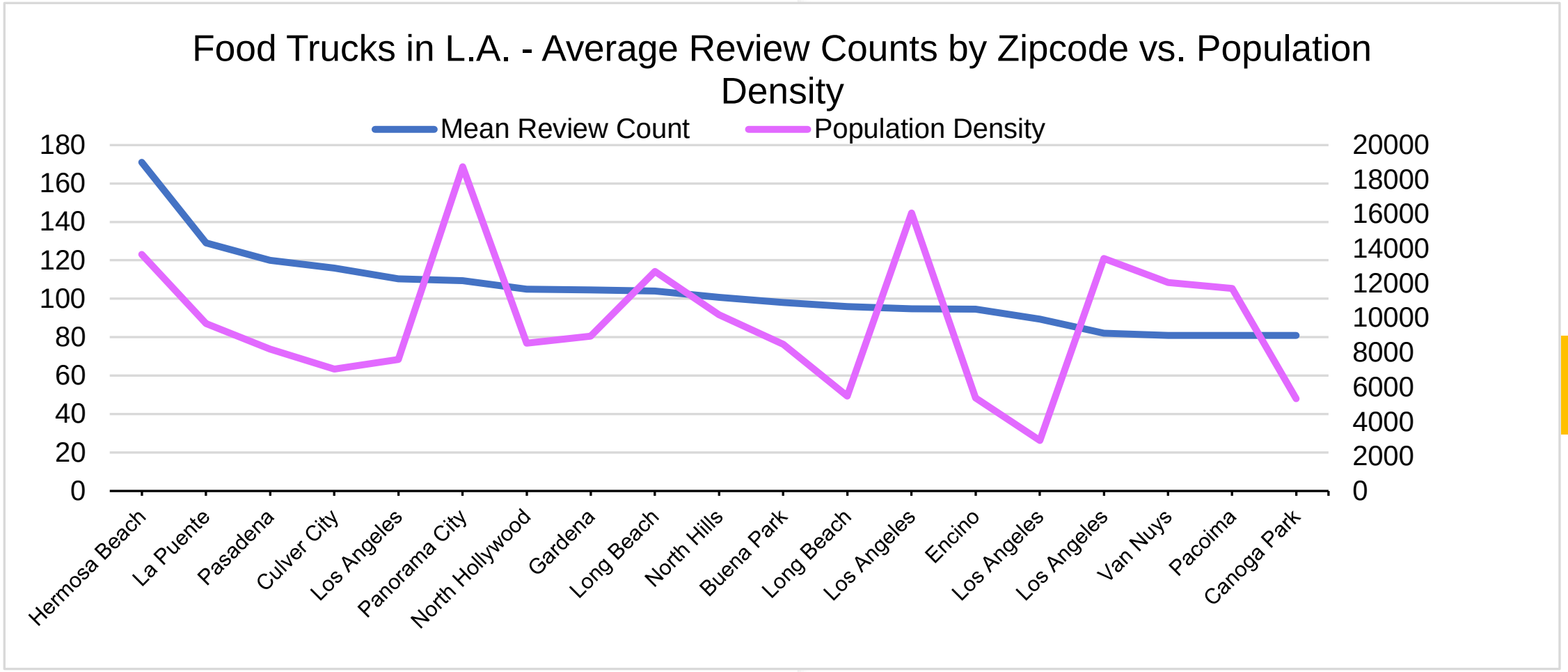


Additional Analysis: Average Review Counts by Zipcode

- *Research question*: **Instead of looking at individual review counts, is there a relationship between zipcode-level average review counts and demographic characteristics of the zipcode?**
- Hypothesis: Average food truck review counts are likely to be correlated with areas that have high disposable income – e.g. high median home values, high median household incomes, and low population density.
- Result: Among the top 20 zipcodes by average review count, there appears to be a clear visual trend between average review counts by Zipcode and median home values, median household incomes, and population density.


Food Trucks in L.A. - Average Review Counts by Zipcode vs. Median Home Value
180
160
140
120
100
80
60
40
20
0
$1,200,000
$1,000,000
$800,000
$600,000
$400,000
$200,000
$-
Mean Review Count
Median Home Value
Hermosa Beach
La Puente
Pasadena
Culver City
Los Angeles
Panorama City
North Hollywood
Gardena
Long Beach
North Hills
Buena Park
Long Beach
Los Angeles
Encino
Los Angeles
Los Angeles
Van Nuys
Pacoima
Canoga Park


Food Trucks in L.A. - Average Review Counts by Zipcode vs. Median Household Income
Mean Review Count
Median Household Income
180
160
140
120
100
80
60
40
20
0
$120,000
$100,000
$80,000
$60,000
$40,000
$20,000
$-
Hermosa Beach
La Puente
Pasadena
Culver City
Los Angeles
Panorama City
North Hollywood
Gardena
Long Beach
North Hills
Buena Park
Long Beach
Los Angeles
Encino
Los Angeles
Los Angeles
Van Nuys
Pacoima
Canoga Park


Food Trucks in L.A. - Average Review Counts by Zipcode vs. Population Density
Mean Review Count
Population Density
180
160
140
120
100
80
60
40
20
0
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Hermosa Beach
La Puente
Pasadena
Culver City
Los Angeles
Panorama City
North Hollywood
Gardena
Long Beach
North Hills
Buena Park
Long Beach
Los Angeles
Encino
Los Angeles
Los Angeles
Van Nuys
Pacoima
Canoga Park

Areas for Further Analysis

- Segment analysis by cuisine type
  - Could use Google Places to attempt to backfill
- Normalize values in current linear regression analysis
- Linear regression on average review counts
  - With additional time, could dive in further to explore statistical significance

Challenges

- Yelp API constraints
  - Nesting issues
  - Cataloging issues (e.g. alias can be both food truck and cuisine type)
  - Searching across multiple cities and integrating results
- Github
  - Some issues collaborating and sharing content