

IDENTIFYING ETHNIC NEIGHBORHOODS OF NEW YORK CITY BY FOOD CULTURE

CAPSTONE PROJECT: THE BATTLE OF NEIGHBORHOODS

YANBO YE

2019-05-20

IDENTIFYING NEIGHBORHOOD ETHNICITY OF NEW YORK CITY IS USEFUL

- New York City is a multicultural city most popular in the United State. The population of New York City is very diverse in ethnicity.
- Most people like to live in a neighborhood with those of their own ethnic group.
- The ethnicity of food venues nearby could be used to cluster neighborhoods using k-means.
- Identifying the dominant race of a neighborhood would be helpful to those who want to live or do bushiness there.

DATA ACQUISITION AND CLEANING

- The neighborhood name and location data of New York City was downloaded from [New York University](#).
- The food venue data with name, category and location of each neighborhood was downloaded from Foursquare's [explore api](#).
- The relation of food venue categories was downloaded from Foursquare's [categories api](#).
- 3805 food venues of 40 categories, which are literally ethnical, were included as feature data for 271 neighborhoods.

NEIGHBORHOOD ETHNICITY WEIGHT MEASUREMENT

- For one neighborhood, the vector of venue count of each categories denote as:

$$\vec{C} = (c_1, c_2, \dots, c_n)$$

- Weight contribution of categories, we'll define neighborhood feature vector as:

$$\vec{F} = \frac{\vec{C}}{\sum_{i=1}^n c_i}$$

- For all m neighborhood, we got our feature matrix as:

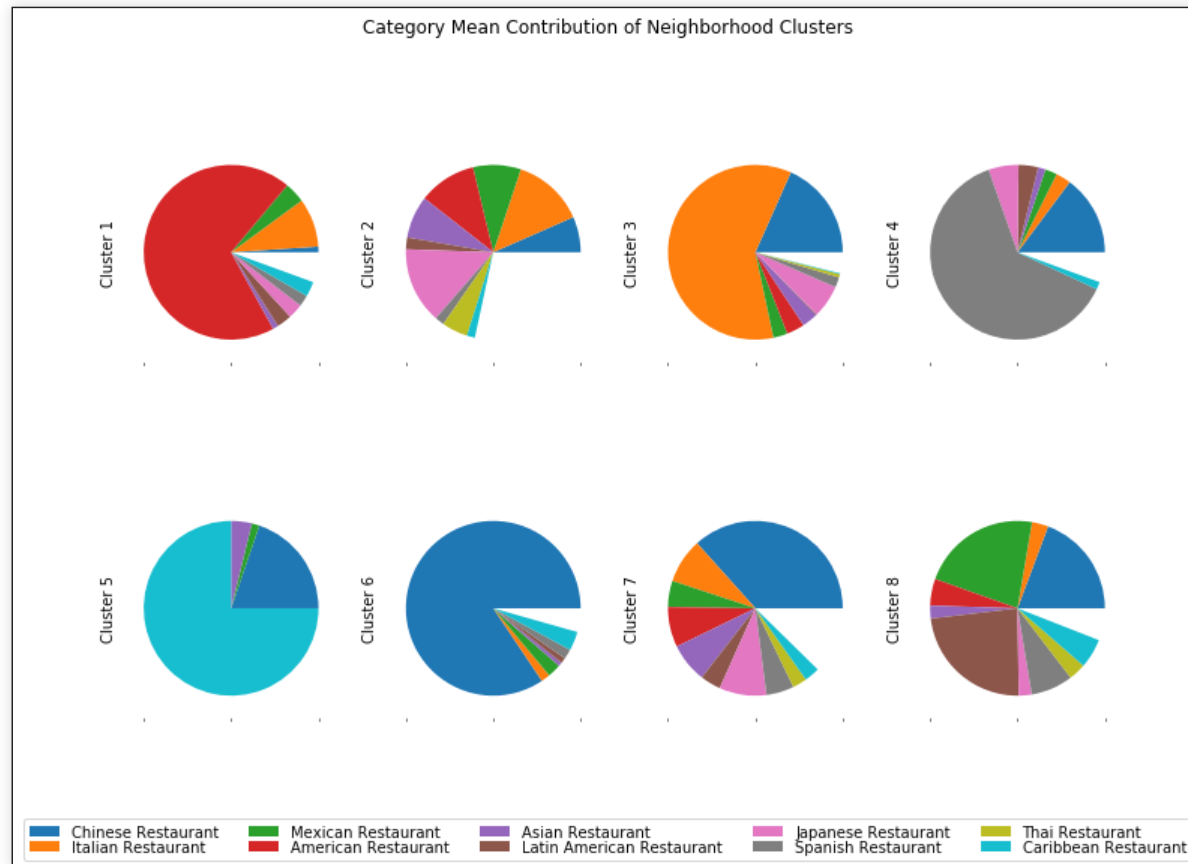
$$F = \begin{pmatrix} \vec{F}_1 \\ \vec{F}_2 \\ \vdots \\ \vec{F}_m \end{pmatrix}$$

CLUSTER ETHNICITY WEIGHT MEASUREMENT

- After clustering, we need to examine the cluster and assign each cluster an label. Suppose we got k clusters. For each cluster with N neighborhoods, we use the mean to measure each category weight. We got our category weight matrix of clusters as:

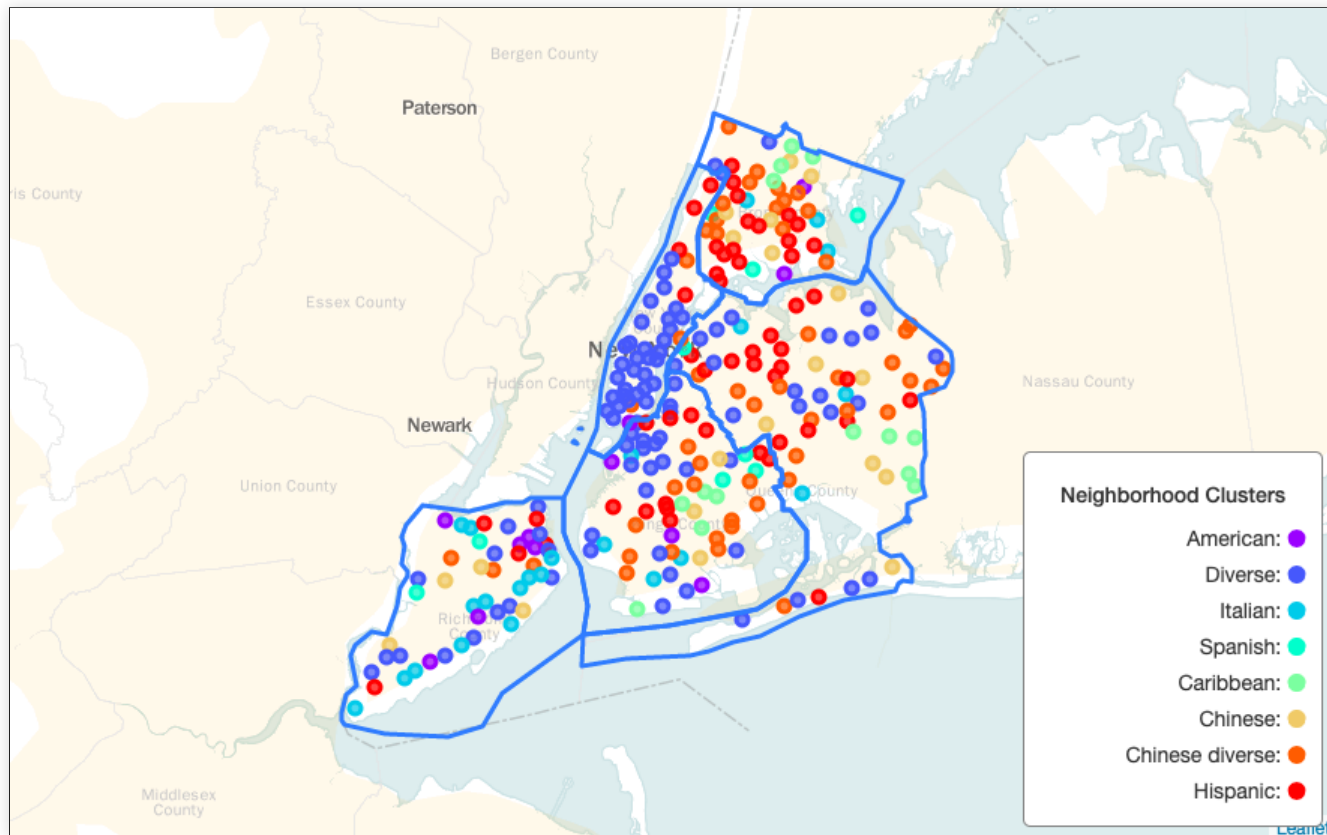
$$E = \begin{pmatrix} \frac{\sum_{i=1}^{N_1} \vec{F}_i}{N_1} \\ \frac{\sum_{i=1}^{N_2} \vec{F}_i}{N_2} \\ \vdots \\ \frac{\sum_{i=1}^{N_k} \vec{F}_i}{N_k} \end{pmatrix}$$

TOP 10 CATEGORY WEIGHTS OF CLUSTERS WITH RAW CATEGORIES

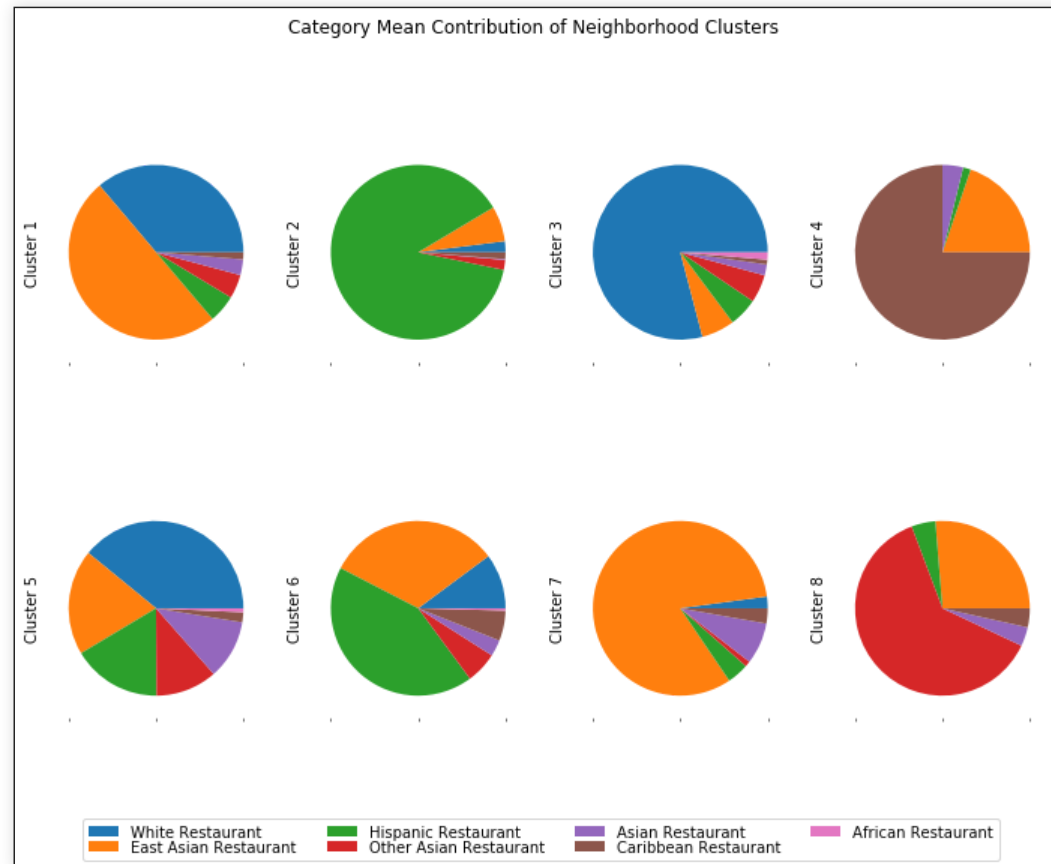


Clusters were labeled as: American, Diverse, Italian, Spanish, Caribbean, Chinese, Chinese diverse, Hispanic.

NEIGHBORHOODS DISTRIBUTION OF DIFFERENT CLUSTERS WITH RAW CATEGORIES

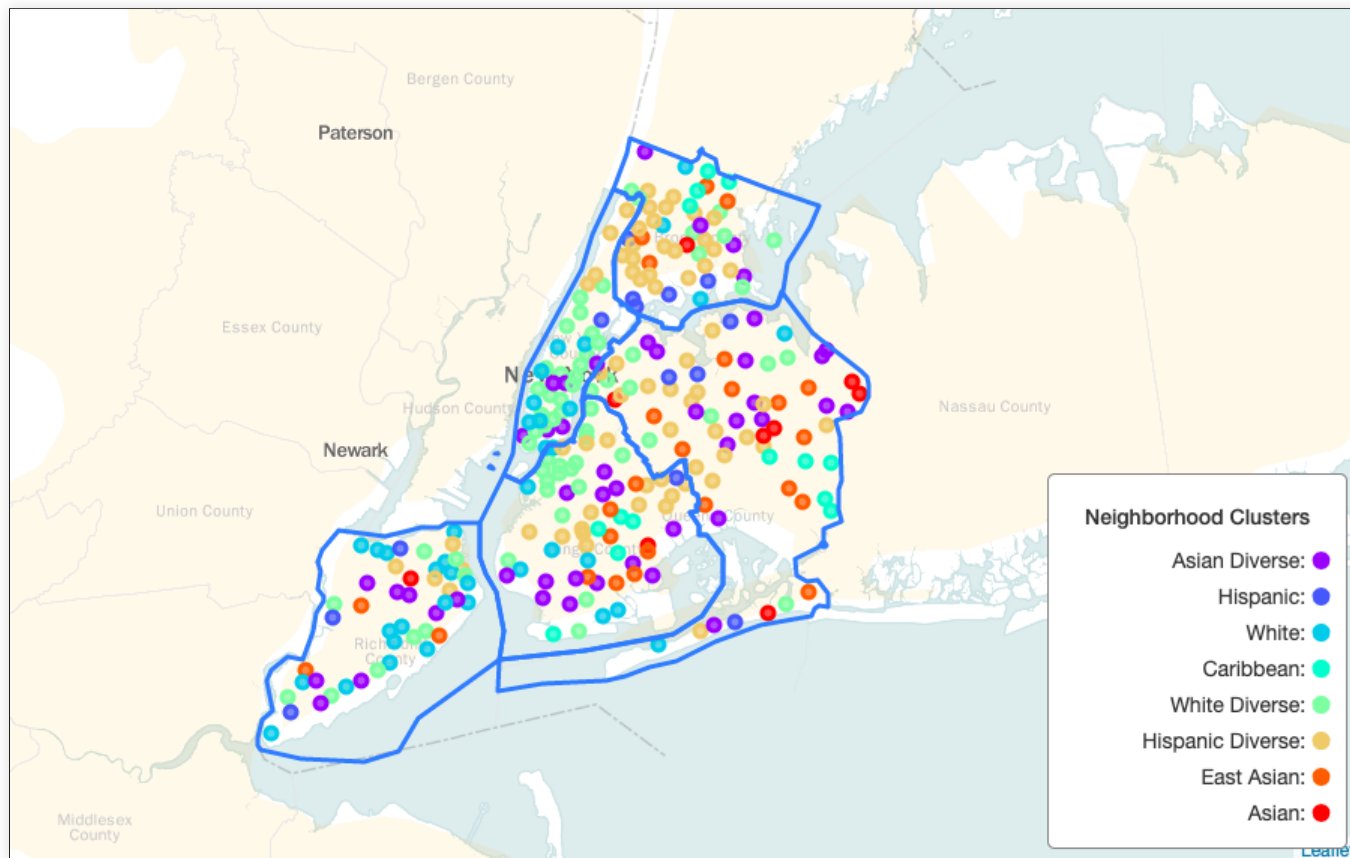


CATEGORY WEIGHTS OF CLUSTERS WITH GENERALIZED CATEGORIES

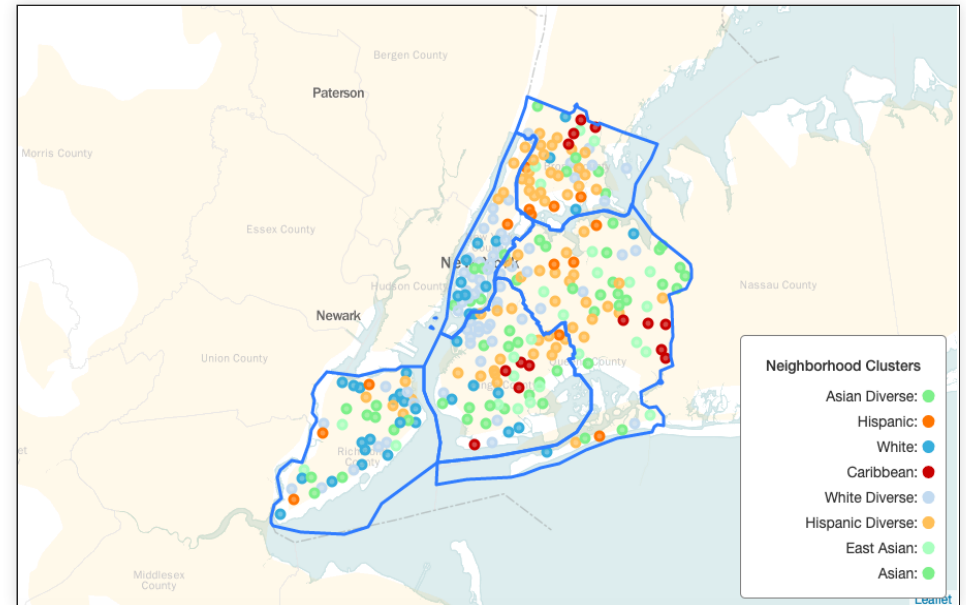
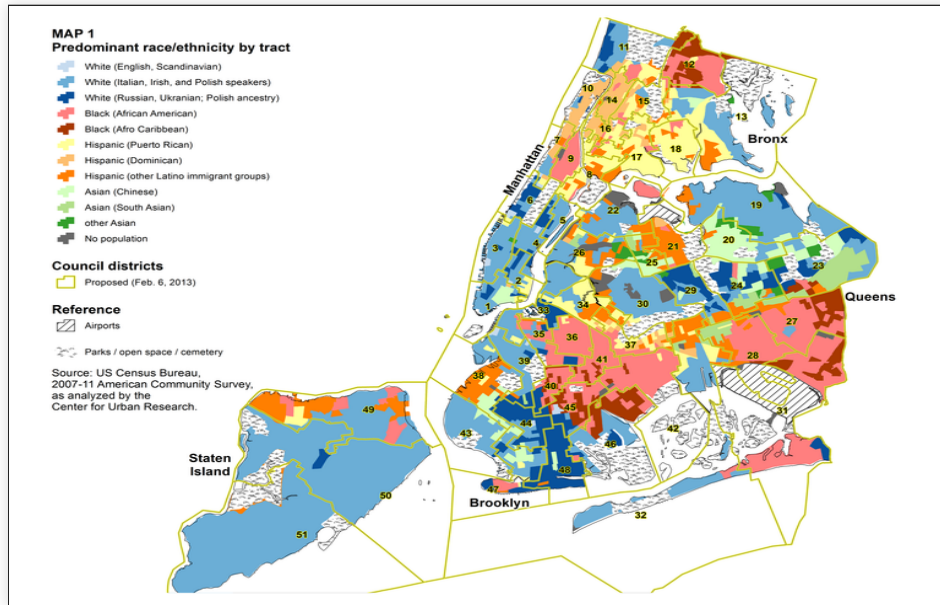


Clusters were labeled as: Asian Diverse, Hispanic, White, Caribbean, White Diverse, Hispanic Diverse, East Asian, Asian.

NEIGHBORHOODS DISTRIBUTION OF DIFFERENT CLUSTERS WITH GENERALIZED CATEGORIES



COMPARISON WITH REAL ETHNICITY DISTRIBUTION DATA OF NEW YORK CITY



Left figure showing the **Predominant race/ethnicity by tract** from a research called **"COMMUNITIES OF INTEREST" IN NEW YORK CITY**.

Right one was redrawn of the second result by taking the same colors of respective ethnic groups.

COMPARISON SIMILARITY AND DIVERSITY

- Similarity:
 1. Neighborhoods in *Staten Island* and *Manhattan* are mostly **White** dominant.
 2. The **Hispanic** neighborhoods are mainly distributed in *The Bronx* and area between *Brooklyn* and *Queens*.
 3. The **Caribbean** neighborhoods can be divided into three geographical clusters in *The Bronx*, *Queens* and *Brooklyn*.
- Diversity:
 1. The **Asian** neighborhoods distribution pattern, which scatter everywhere, is not like the previous one that most **Asian** neighborhoods are in *Queens*.
 2. The **African** neighborhoods are not shown on our map.

CONCLUSION

- We clustered neighborhoods in New York City into different ethnic groups based on the food venue data.
- The result is comparable to the real race distribution of New York City.
- Thus we provide a good way to cluster and identify the dominant ethnicity of neighborhoods.
- Include more categories by study the culture background of all food categories might improve our result.