

IDENTIFYING FOOD AND HEALTHCARE DESERTS AND THEIR POTENTIAL IMPACT

Capstone Project – IBM Data Science Certification on Coursera

Brad Mills

- ❑ Geographic areas where a resident's access to affordable, healthy food options is limited or nonexistent due to the absence of grocery stores within convenient travelling distance are known as **Food deserts**.¹
- ❑ **Healthcare deserts** are areas where there is inadequate access to one or more kinds of medical services.²

This capstone projects looks at automated ways of identifying Food and Healthcare deserts using location data and examines their potential impact by assessing correlations with the graduation rates of neighboring schools.

Location data was obtained via the Foursquare API (*foursquare.com*)

Food and Healthcare deserts were classified based on the types of businesses in the neighboring areas surrounding Indianapolis area schools

Schools with similar neighborhood features were then grouped using three different methods:

- A simple categorization scheme based on the total count of related businesses
- An unsupervised clustering algorithm using this same total business count data
- The same unsupervised clustering algorithm using more detailed business counts

These groupings/clusters of schools were then linked to data from the Indiana Dept of Education³ and analyzed to see if they shared similar graduation rates

LINKED DATA SNAPSHOT

Examples of a JSON feed from Foursquare, Indiana DOE spreadsheet data, and the linked data after automated cleaning and name standardization

Corp Id	CORPORATION NAME	School Id	SCHOOL NAME	Cohort N	Grad N	2019 State Grad Rate	Non-Waiver Grad N	2019 State Non-Waiver Grad Rate	
0	15	Adams Central Community Schools	21	Adams Central High School	99	98	0.9899	95	0.9596
1	25	North Adams Community Schools	29	Bellmont Senior High School	173	160	0.9249	146	0.8439
2	35	South Adams Schools	23	South Adams High School	73	63	0.863	56	0.7671
3	125	M S D Southwest Allen County Schls	47	Homestead Senior High School	578	547	0.9464	528	0.9135

```
{
  'meta': {
    'code': 200,
    'requestId': '5f1e32952321bc05cdfc1a42'
  },
  'response': {
    'headerLocation': 'Indianapolis',
    'headerFullLocation': 'Indianapolis',
    'headerLocationGranularity': 'city',
    'query': 'high school',
    'totalResults': 101,
    'suggestedBounds': {
      'ne': {
        'lat': 39.98433331600022,
        'lng': -85.8778577624528
      },
      'sw': {
        'lat': 39.55233288399978,
        'lng': -86.4388426375472
      }
    },
    'groups': [
      {
        'type': 'Recommended Places',
        'name': 'recommended',
        'items': [
          {
            'reasons': {
              'count': 0,
              'items': [
                {
                  'summary': 'This spot is popular',
                  'type': 'general',
                  'reasonName': 'globalInteractionReason'
                }
              ]
            },
            'venue': {
              'id': '4b476ee4f964a520e93126e3',
              'name': 'Decatur Central High School',
              'location': {
                'address': '5251 Kentucky Ave',
                'crossStreet': 'at Gatwick Dr',
                'lat': 39.68781376487072,
                'lng': -86.27571795903808,
                'labeledLatLngs': [
                  {
                    'label': 'display',
                    'lat': 39.68781376487072,
                    'lng': -86.27571795903808
                  }
                ],
                'distance': 13465,
                'postalCode': '46221',
                'cc': 'US',
                'city': 'Indianapolis',
                'state': 'IN',
                'country': 'United States',
                'formattedAddress': [
                  '5251 Kentucky Ave (at Gatwick Dr)',
                  'Indianapolis, IN 46221',
                  'United States'
                ]
              },
              'categories': [
                {
                  'id': '4bf58dd8d48988d13d941735',
                  'name': 'High School',
                  'pluralName': 'High Schools',
                  'shortName': 'High School',
                  'icon': {
                    'prefix': 'https://ss3.4sqi.net/img/categories_v2/building/school_',
                    'suffix': '.png',
                    'primary': True
                  },
                  'photos': {
                    'count': 0,
                    'groups': []
                  },
                  'referralId': 'e-0-4b476ee4f964a520e93126e3-0'
                }
              ]
            }
          }
        ]
      }
    ]
  }
}
```

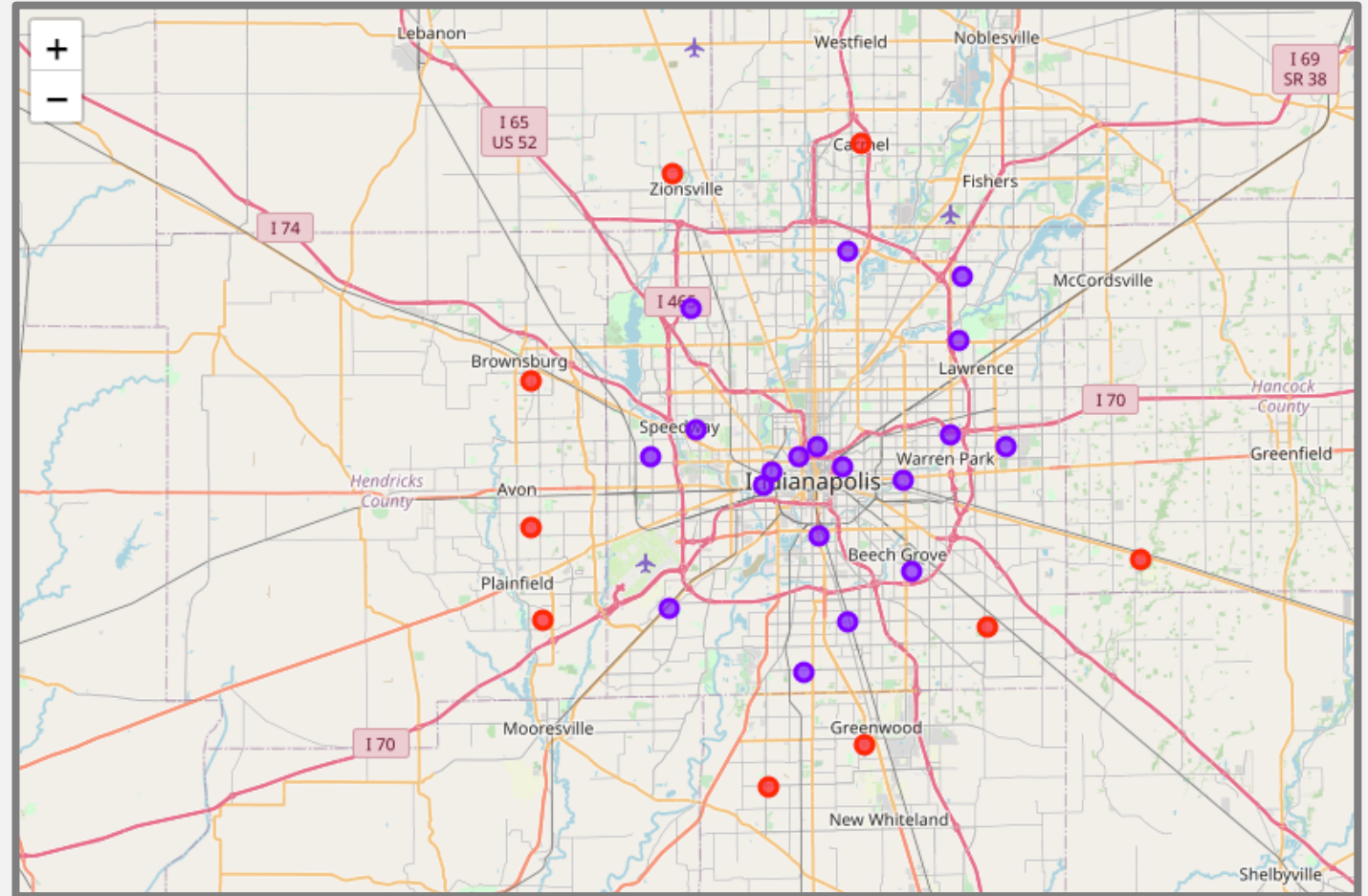
	School	Latitude	Longitude	Setting	GradRate	ClassSize	Corporation
0	Decatur Central	39.687814	-86.275718	Urban	0.7257	412	M S D Decatur Township
1	Greenwood	39.602884	-86.116524	Rural	0.8611	288	Greenwood Community Sch Corp
2	Crispus Attucks	39.782697	-86.169893	Urban	0.8685	213	Indianapolis Public Schools
3	Arsenal Technical	39.776575	-86.133930	Urban	0.7736	605	Indianapolis Public Schools
4	Herron	39.789116	-86.155428	Urban	0.9474	209	Herron Charter

MAP OF LOCATED AREA SCHOOLS

5

Map generated in Python
identifying Indianapolis area
Public schools

Color of marker
determines if school
should be considered
Urban or **Rural** to account
for travel distances and
modes of transportation



IDENTIFYING HEALTHCARE DESERTS

6

A Foursquare Location query
result converted to a pandas
dataframe

	Neighborhood	Setting	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Decatur Central	Urban	39.687814	-86.275718	Wesley G. Lackey MD-Midwest Center for Joint R...	39.687837	-86.280184	Doctor's Office
1	Decatur Central	Urban	39.687814	-86.275718	Richard W. Jackson MD-Midwest Center for Joint...	39.687771	-86.280205	Doctor's Office
2	Decatur Central	Urban	39.687814	-86.275718	Midwest Specialty Surgery Center	39.687575	-86.281323	Hospital
3	Decatur Central	Urban	39.687814	-86.275718	Joshua L. Carter MD-Midwest Center for Joint R...	39.687710	-86.281662	Doctor's Office
4	Decatur Central	Urban	39.687814	-86.275718	Michael E. Berend MD-Midwest Center for Joint ...	39.687710	-86.281662	Doctor's Office

	Neighborhood	Doctor's Office	Emergency Room	Hospital	Mental Health Office	Pharmacy	Urgent Care Center	Total
0	Arsenal Technical	13	0	0	0	0	0	13
1	Avon	18	0	1	0	3	0	22
2	Ben Davis	0	1	1	0	2	0	4
3	Brownsburg	15	0	0	0	3	1	19
4	Carmel	86	1	5	2	6	0	100
5	Center Grove	2	0	1	0	0	0	3
6	Crispus Attucks	76	2	11	4	1	0	94
7	Decatur Central	6	0	1	0	0	0	7

IDENTIFYING FOOD DESERTS

7

A Foursquare Location query result converted to a pandas dataframe

	Neighborhood	Setting	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Decatur Central	Urban	39.687814	-86.275718	Los Patios Mexican Store	39.694170	-86.269020	Grocery Store
1	Greenwood	Rural	39.602884	-86.116524	Aldi	39.594531	-86.101949	Supermarket
2	Greenwood	Rural	39.602884	-86.116524	Gordon Food Service	39.627259	-86.121817	Grocery Store
3	Greenwood	Rural	39.602884	-86.116524	Kroger	39.584052	-86.102857	Supermarket
4	Greenwood	Rural	39.602884	-86.116524	Fresh Thyme Farmers Market	39.638519	-86.130807	Grocery Store

	Neighborhood	Farmers Market	Grocery Store	Organic Grocery	Supermarket	Total
0	Arsenal Technical	2	0	0	0	2
1	Avon	1	4	1	8	14
2	Brownsburg	0	1	1	2	4
3	Carmel	2	8	1	4	15
4	Center Grove	0	3	0	3	6
5	Decatur Central	0	1	0	0	1

CATEGORIZING RESOURCE DESERTS

8

First grouping method:

Neighborhoods categorized as High, Medium, or Low Risk based on Total count of grocery or medical facilities

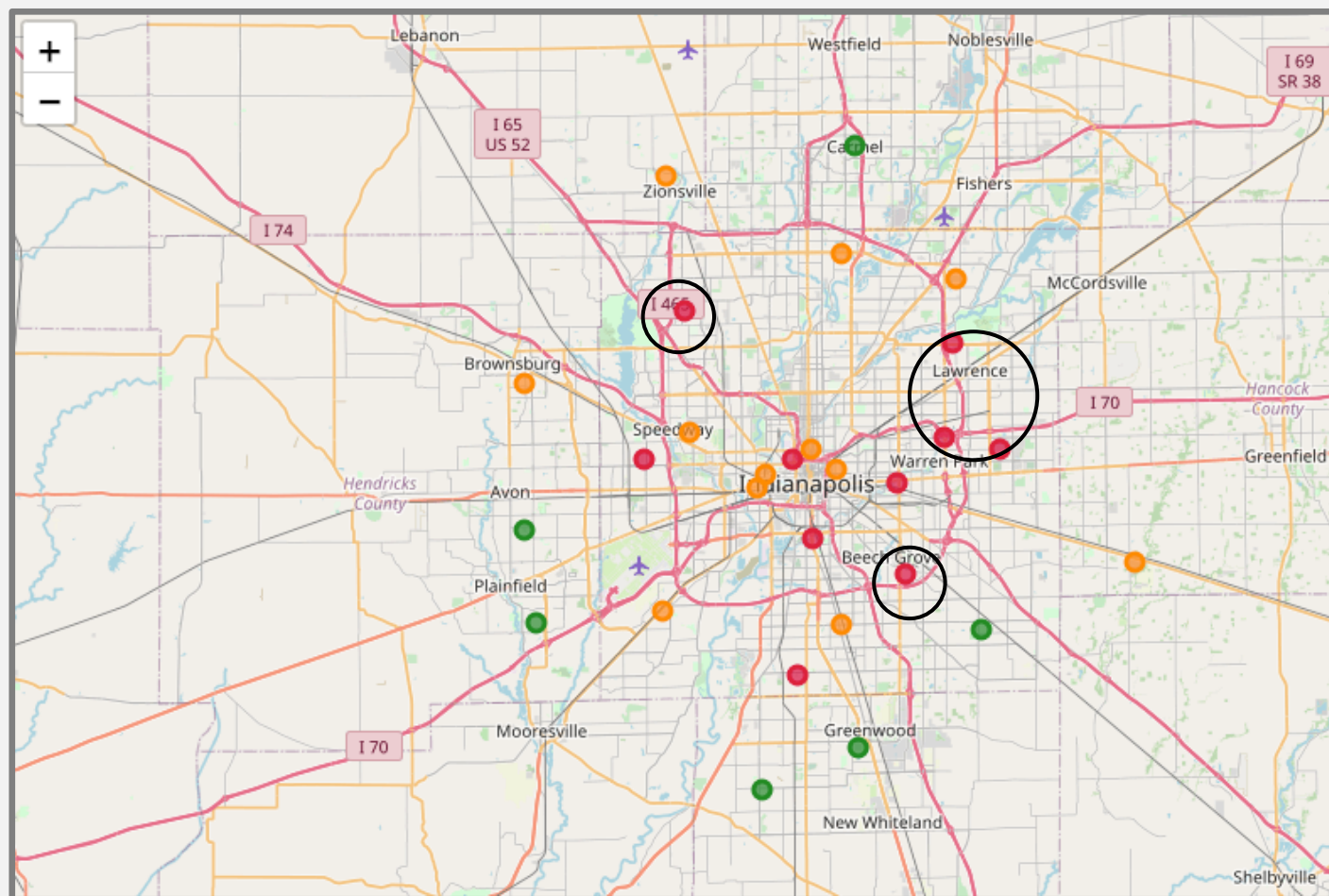
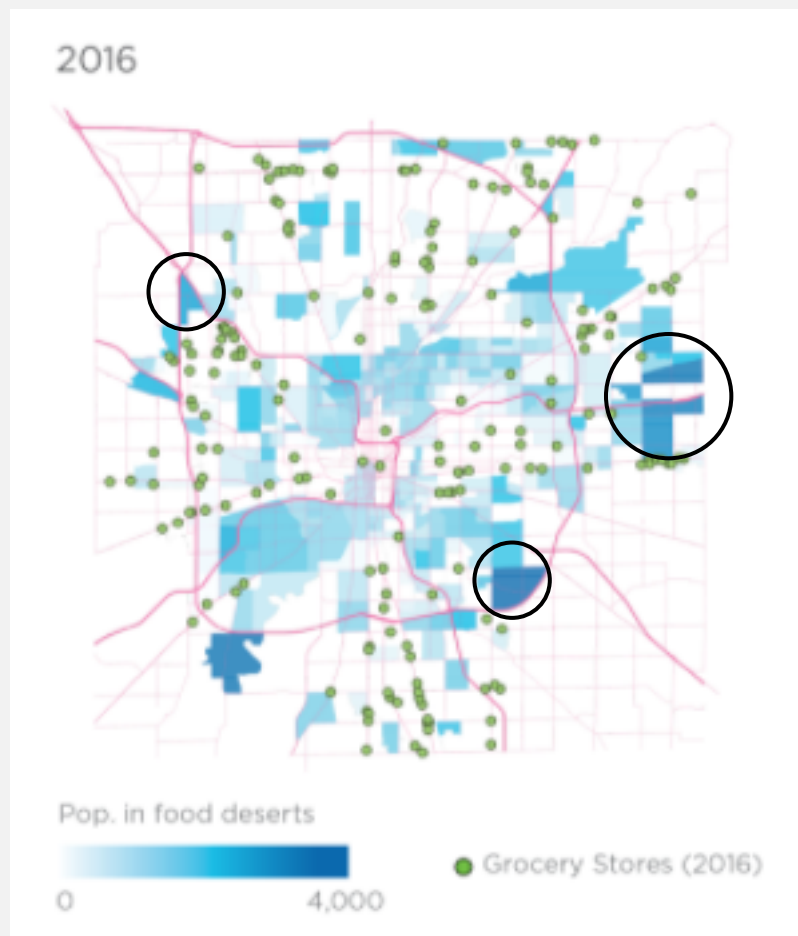
	Neighborhood	Farmers Market	Grocery Store	Organic Grocery	Supermarket	Total	Food Desert Risk Indicator
0	Arsenal Technical	2	0	0	0	2	Medium
1	Avon	1	4	1	8	14	Low
2	Beech Grove	0	0	0	0	0	High
3	Ben Davis	0	0	0	0	0	High
4	Brownsburg	0	1	1	2	4	Medium
5	Carmel	2	8	1	4	15	Low
6	Center Grove	0	3	0	3	6	Low
7	Crispus Attucks	0	0	0	0	0	High
8	Decatur Central	0	1	0	0	1	Medium
9	Emmerich Manual	0	0	0	0	0	High

	Neighborhood	Doctor's Office	Emergency Room	Hospital	Mental Health Office	Pharmacy	Urgent Care Center	Total	Medical Desert Risk Indicator
0	Arsenal Technical	13	0	0	0	0	0	13	Medium
1	Avon	18	0	1	0	3	0	22	Medium
2	Beech Grove	0	0	0	0	0	0	0	High
3	Ben Davis	0	1	1	0	2	0	4	High
4	Brownsburg	15	0	0	0	3	1	19	Medium
5	Carmel	86	1	5	2	6	0	100	Low
6	Center Grove	2	0	1	0	0	0	3	High
7	Crispus Attucks	76	2	11	4	1	0	94	Low
8	Decatur Central	6	0	1	0	0	0	7	High
9	Emmerich Manual	3	0	0	0	0	0	3	High

FOOD DESERT MAPS

9

Similar **High** Risk areas identified
as other published reports⁵

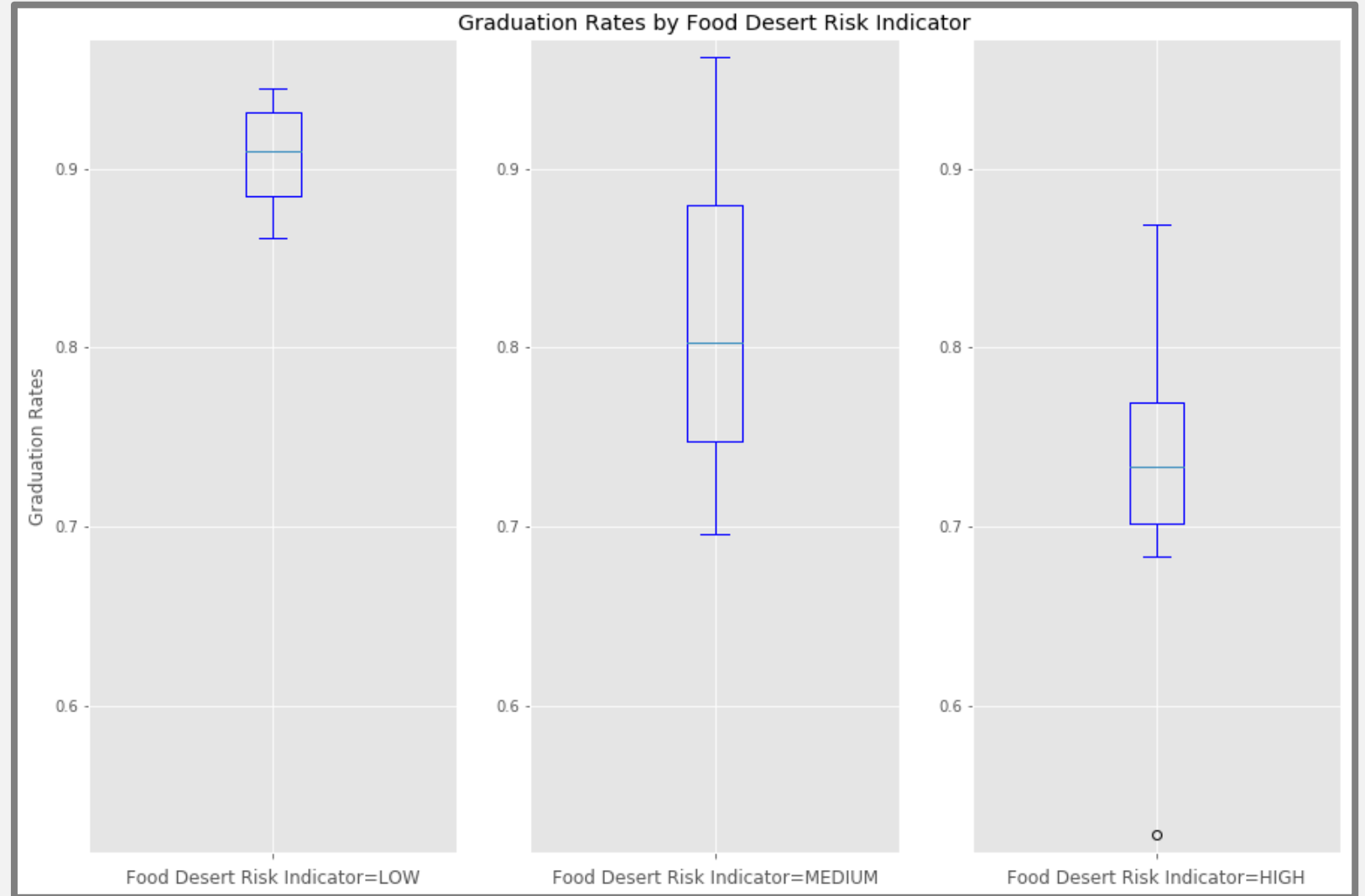


ASSESSING DATA CORRELATIONS

10

Strong visual indication that the mean **Graduation Rate** for neighborhood groupings decreases as the risk of losing local **Food resources** increases

Area school neighborhoods were grouped by their Total count of readily available **Food resources**

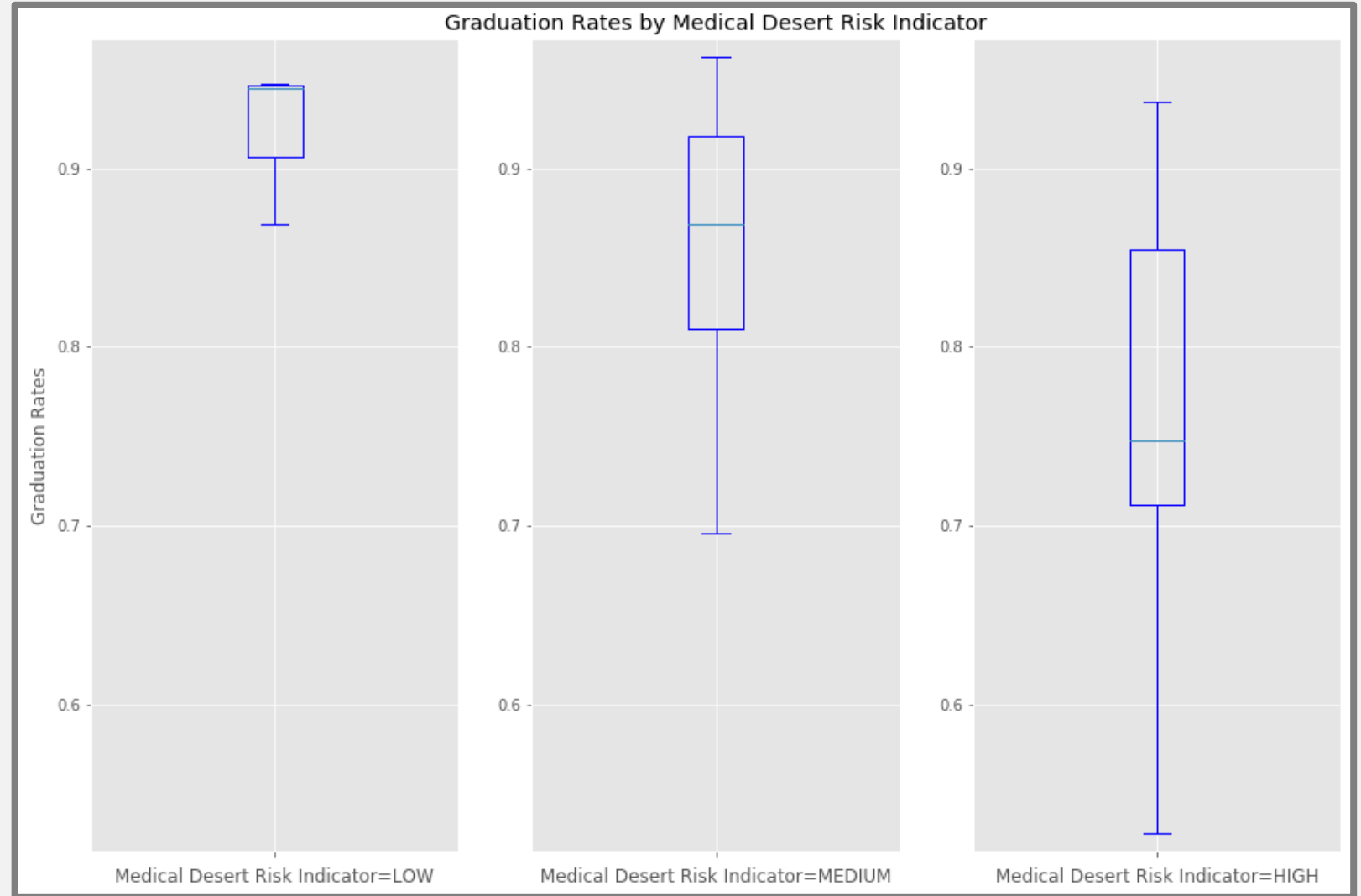


ASSESSING DATA CORRELATIONS

11

Visual indication that the mean **Graduation Rate** for neighborhood groupings decreases as the risk of losing local **Health resources** increases

Area school neighborhoods were grouped by their Total count of readily available **Medical resources**



FOOD DESERT CATEGORY STATS

12

ANOVA analysis verifying that the observed difference in mean **Graduation Rates** among the neighborhood groupings by **Food Resource** risk is statistically significant at $p < .01$

Risk=Low		Risk=Medium		Risk=High	
count	6.000000	count	12.000000	count	10.000000
mean	0.906700	mean	0.819283	mean	0.733490
std	0.032789	std	0.093472	std	0.095074
Source of variability	SS	df	Mean Square	F Ratio	
Between groups	0.1155	2	0.0577	7.90	
Within groups	0.1828	25	0.0073		
Total	0.2983	27			

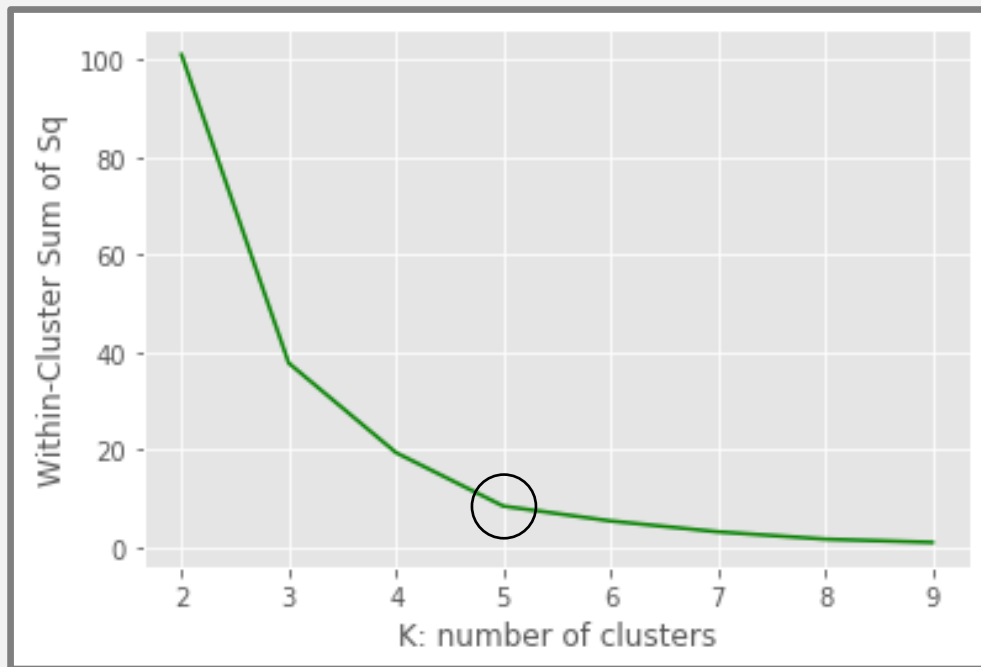
Analysis of variance (ANOVA) indicates that there are significant differences among the 3 groups, $F(2, 25) = 7.90, p < .01$

[Critical value: $F_{.01} = 5.57$]

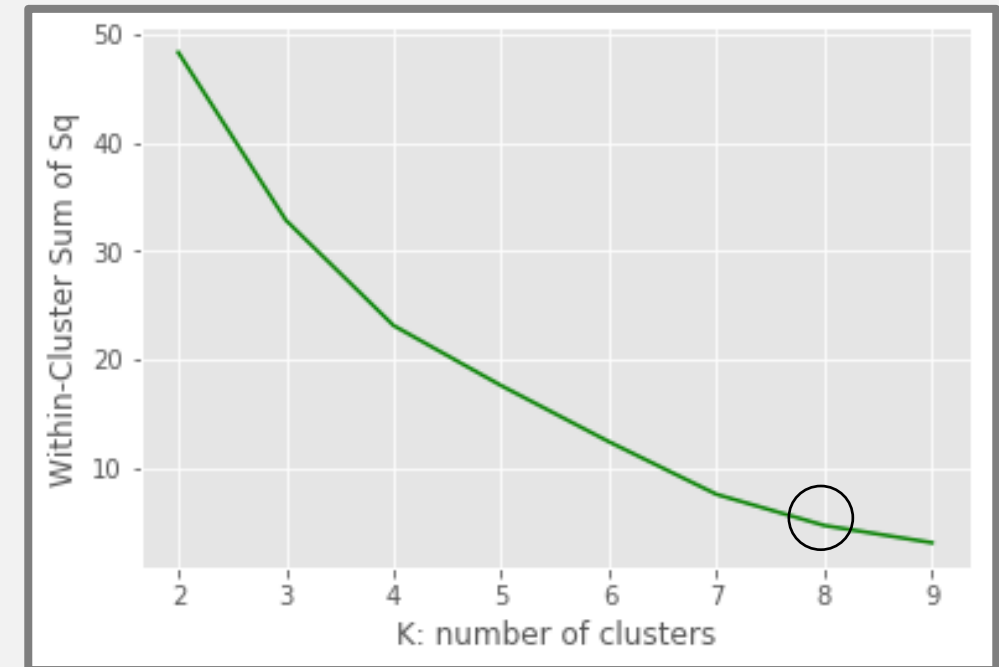
PICKING K FOR K-MEANS CLUSTERING

13

Looking for inflection points, aka the “elbow” heuristic, to pick a logical K value to use for the number of clusters in K-means



Method 1: Total resource count



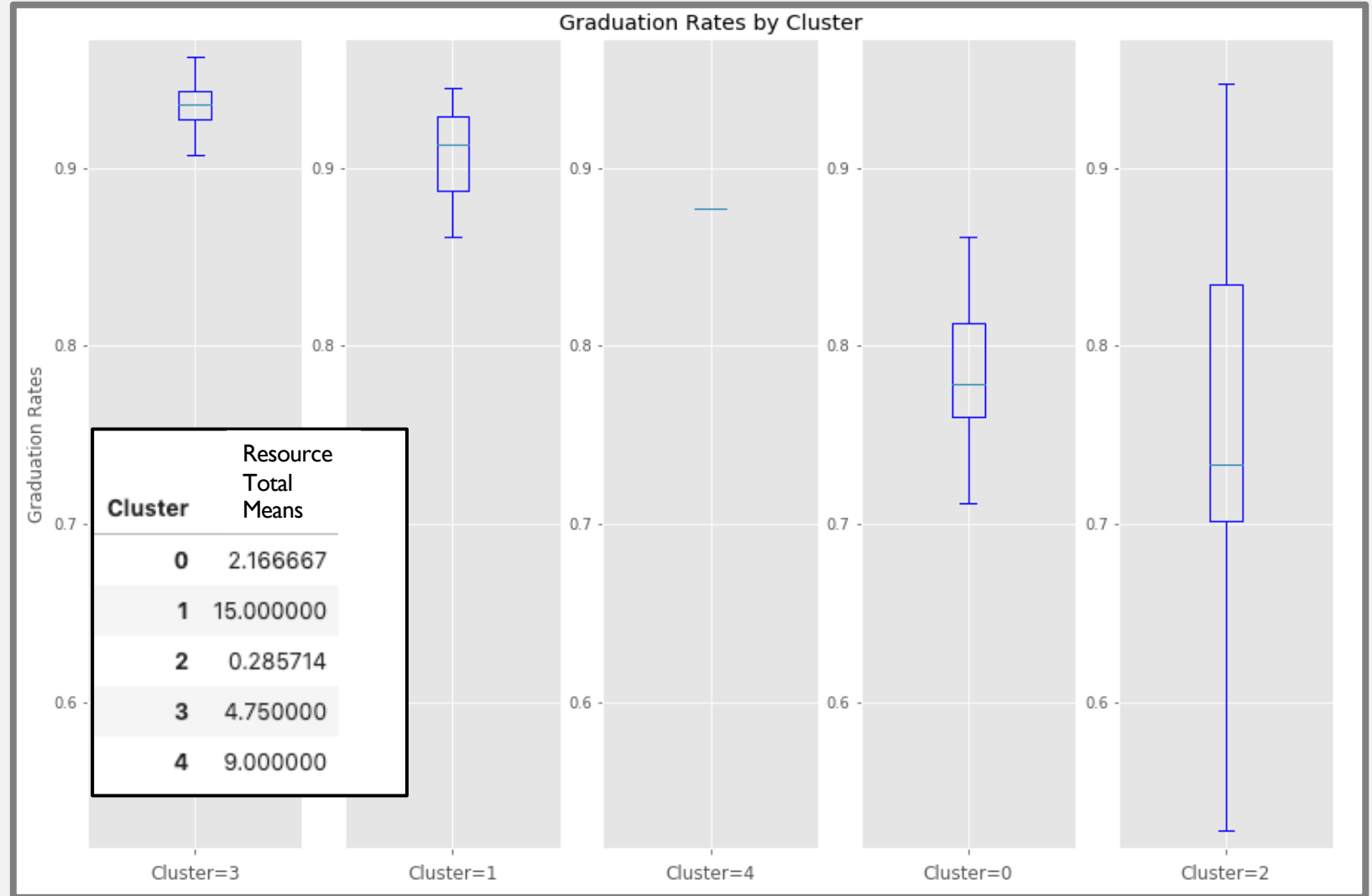
Method 2: Grocery subcategories

K-MEANS CLUSTERING : METHOD I

14

Visual indication that the mean **Graduation Rate** for K-means clusters decreases as the risk of losing local **Food resources** increases

Clusters are based on **Total count** of readily available Food resources in the neighborhood



K-MEANS CLUSTERING : METHOD 2

15

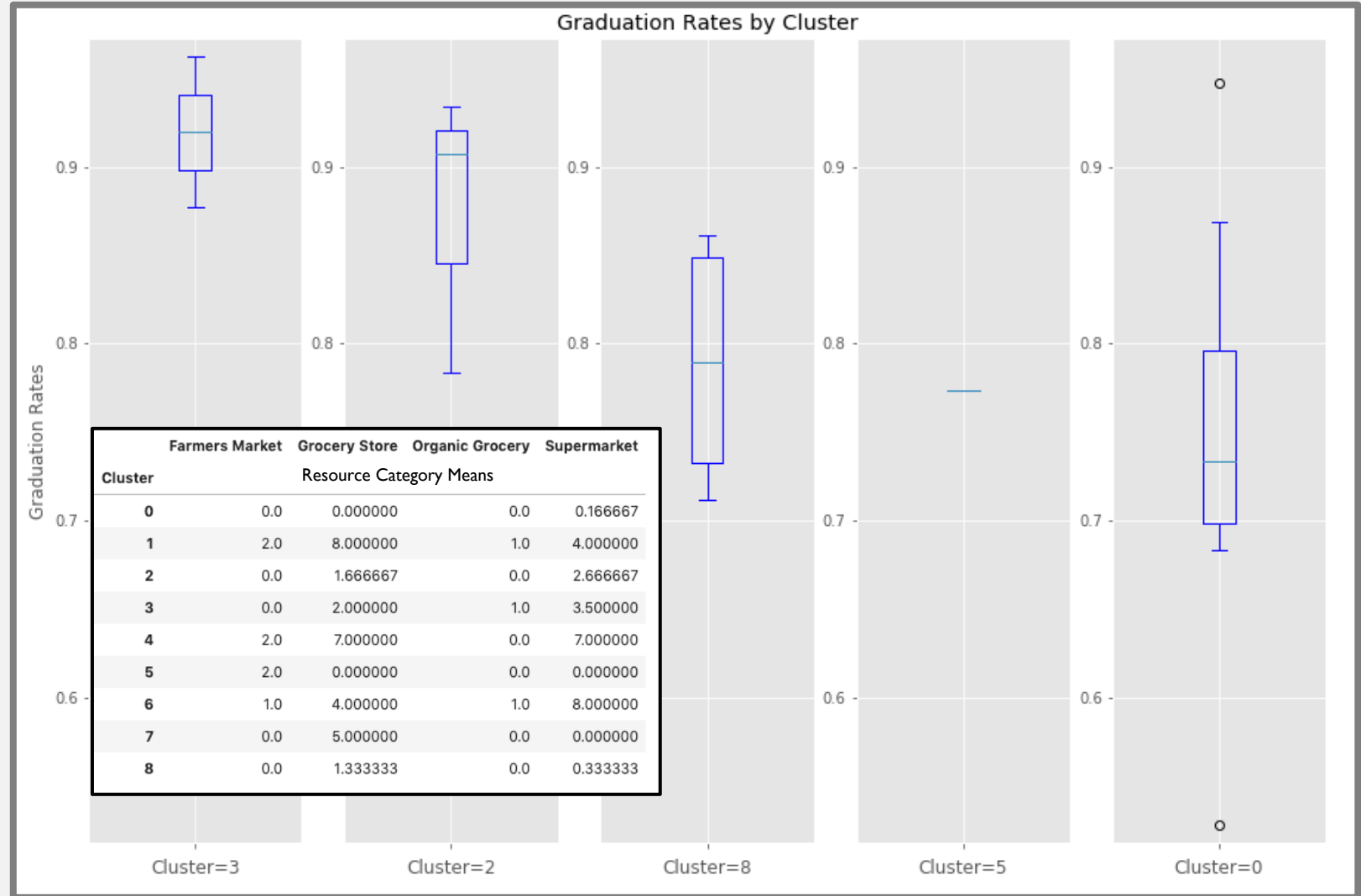
Visual indication that the mean **Graduation Rate** for K-means clusters decreases as the risk of losing local **Food resources** increases

Clusters are based on the **grocery-related subcategories** available in the neighborhood

Interpreting Resource Category Means:

Neighborhoods in Cluster 2, for example, typically have 1-2 Grocery Stores and 2-3 Supermarkets in their immediate area

Cluster 5 is a single neighborhood that only has seasonal Farmers Markets



K-MEANS CLUSTERING : METHOD 2

16

Additional details on an example cluster, Cluster 8,
from clustering neighborhoods on **Grocery subcategories**
and then linking to **Graduation Rates** for analysis

```
count    6.000000
mean     0.789217
std      0.066490
min      0.712200
25%      0.733200
50%      0.789150
75%      0.848850
max      0.861500
Name: Graduation Rate
```

	Neighborhood	Farmers Market	Grocery Store	Organic Grocery	Supermarket	Total	Graduation Rate	Cluster
8	Decatur Central	0	1	0	0	1	0.7257	8
11	George Washington	0	2	0	0	2	0.7557	8
14	Indianapolis Metropolitan	0	2	0	0	2	0.8226	8
17	New Palestine	0	1	0	0	1	0.8576	8
23	Southport	0	1	0	1	2	0.7122	8
24	Speedway	0	1	0	1	2	0.8615	8

REFERENCES

1. <https://foodispower.org/access-health/food-deserts/>
2. https://en.wikipedia.org/wiki/Medical_deserts_in_the_United_States
3. <https://www.doe.in.gov/sites/default/files/accountability/2019-state-grad-rate-data-20191231.xlsx>
4. http://indyfoodcouncil.org/wp-content/uploads/2015/09/Food_Deserts_w_Groc.pdf
5. http://www.savi.org/feature_report/getting-groceries-food-access-across-groups-neighborhoods-and-time/