Possible factors:

① social causes ② where do you live ③ Have access to walking distance amenities? ④ unhealthy lifestyles?

use large scale social determinant of health data, use modern machine learning methods to unravel SDOH and obesity.

Mixture of micro and macro factors

- For a large collection of people, monitor where they move from day to day
- When they go to the grocery store, do they walk?
- How close do they park to the store?

- Do they take the stairs?

- Are there even accessible stairs?

BMI, Income, education attainment, unemployment rates, house quality, access to healthcare, food environment data.

Data

CDC 500 Cities dates

ACS and US department of Agriculture estimates

SDOH

13 features or SD

outcome: obesity

1-25, 25-35, 235

obesity clusters

Getis-Ord Gi* statistics with first order queen contiguity

apply false discovery rate correction → test → spatial dependence

↓

geospatial clustering hot spots of adult prevalence

Regression Modeling

Data Wrangling

scale features ⟶ mean of 0 and SD of 1

reduce the heterogeneity

model selection

"forward and backward" stepwise regression

AIC

VIF 20

Model

Ordinary least squares regression model

adjusted R^2 and AIC select models

Koenker-Bassett test ⟶ heteroskedasticity of random error terms

Jarque-Bera test ⟶ normality of error distribution

Robust Lagrange Multiplier (error) and Robust Lagrange Multiplier (lag) methods ⟶ independence of terms

Order queen contiguity weights ⟶ spatial testing

↓

if dependence was found ⟶ incorporated terms accounted for autocorrelation in the model

spatial lag or spatial error model (SEM)

Grouping analysis

hierarchical clustering unsupervised machine learning

algorithem ——> dependent variable and significantly

associated SDOH across the region

Lack of Physical Activity, obesity, and SDOH

Explored the geographical distribution of lack of

physical activity, obesity, and the top fours features

Visualization and Tools

ArcGIS Pro software ——> spatial distribution

R studio GeoDa software ——> statistical analysis.