Does physical inactivity level and access to healthy food affect obesity rate at the county level?

Helen Peng, Anvith Thumma, Jordan Faustin

Significance

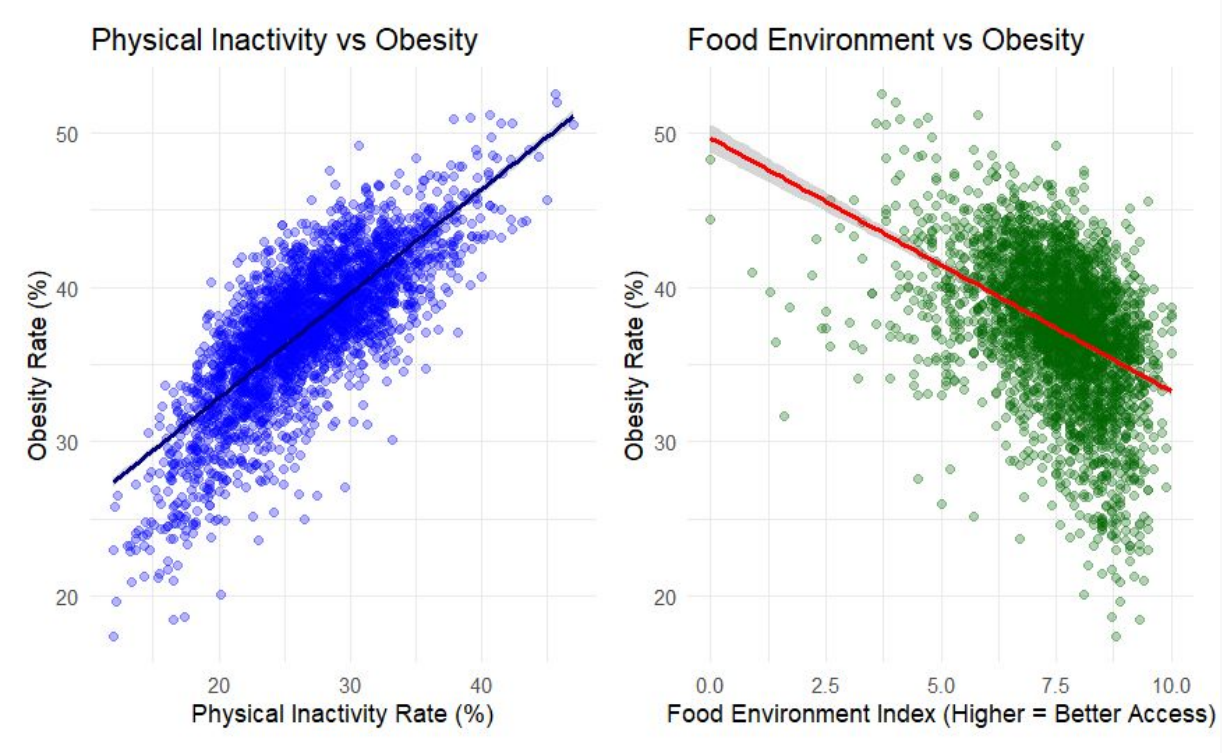
Public Health Crisis: Obesity is a major contributor to chronic diseases (diabetes, heart disease, cancer) and rising healthcare costs.
Disparities in Obesity: Obesity rates vary widely by region, income, and race/ethnicity. Identifying structural barriers (like food access and inactivity) can inform equitable policies.

Data Overview

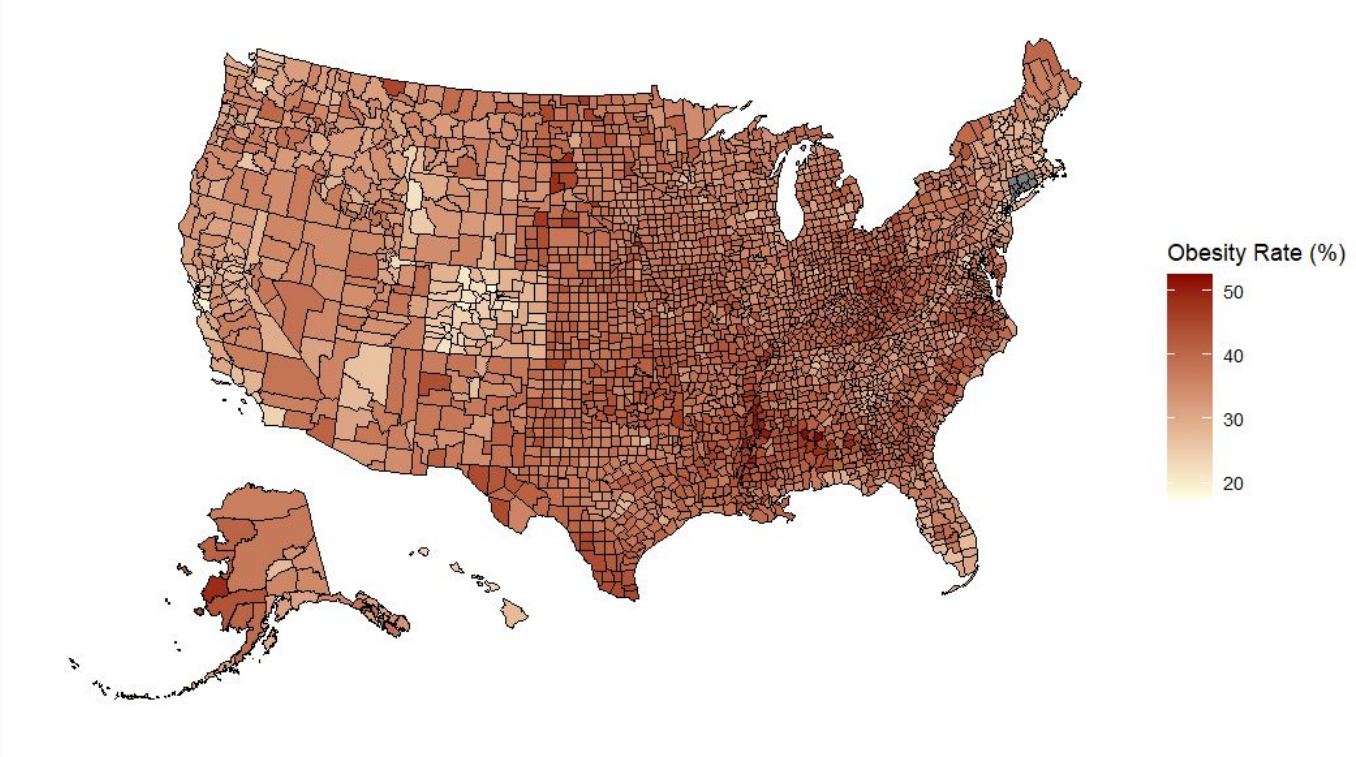
- Observations: Data at the county level across all 50 U.S. states in 2024

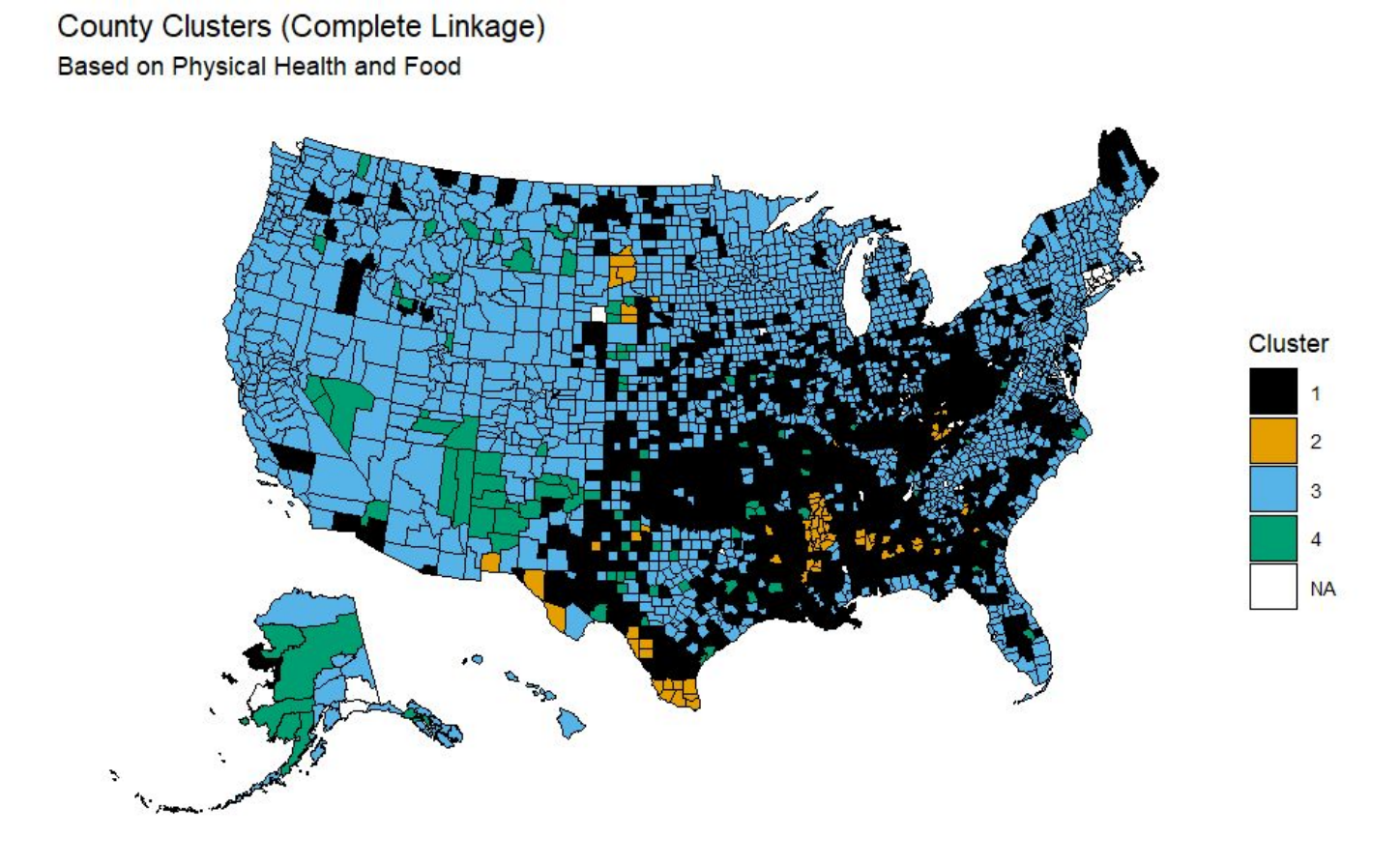
- Variables of Interest:
  - % Adults with Obesity
  - % Physically Inactive
  - Food Environment Index (1-10) - Higher is Better
  - Population
  - State and County Identifiers

Obesity rates increases as physical inactivity increases and better access to food decreases



East coast has higher obesity rate than the west coast




County Clusters (Complete Linkage)
Based on Physical Health and Food
Cluster
1
2
3
4
NA

Cluster 1: High inactivity, decent food access, high obesity
Cluster 2: Very inactive, poor food access, highest obesity
Cluster 3: Most active, best food access, lowest obesity
Cluster 4: Moderately active, poor food access, mid obesity

Cluster Summary: Average Health and Food Metrics by Cluster

Cluster	Number of Counties	Physically Inactive (%)	Food Environment Index	Obesity Rate (%)
1	1322	30.52	7.22	40.44
2	70	38.90	4.49	45.83
3	1618	22.94	8.12	34.45
4	98	27.35	4.51	37.64

Plan of Action

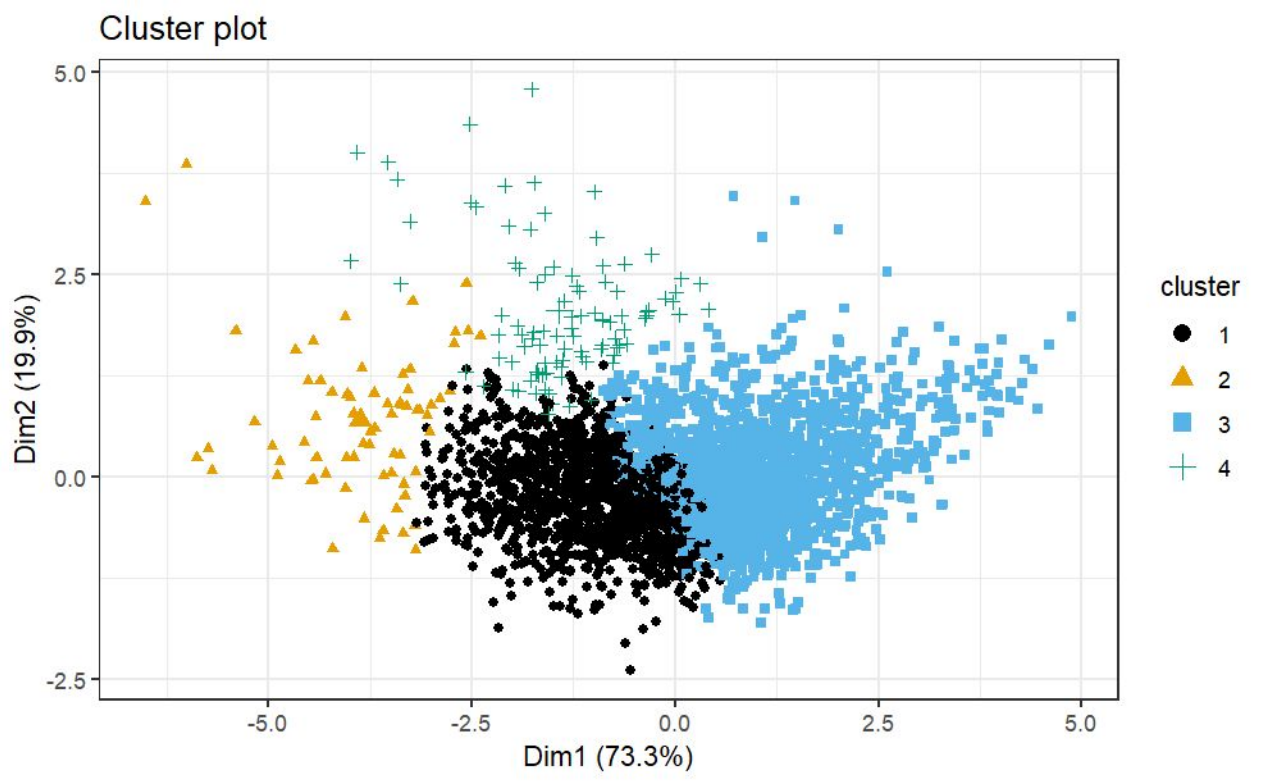
- Conduct EDA on 2024 data, specifically:
  - Create a map to see geographic data of obesity rates
  - Create scatter plots to see the correlation between obesity, food access, and physical inactivity
  - Clustering analysis
- Predictive Modeling (see next slide)
- Evaluate model performance using MSE and other evaluation metrics

Early Thoughts on Modeling Strategy

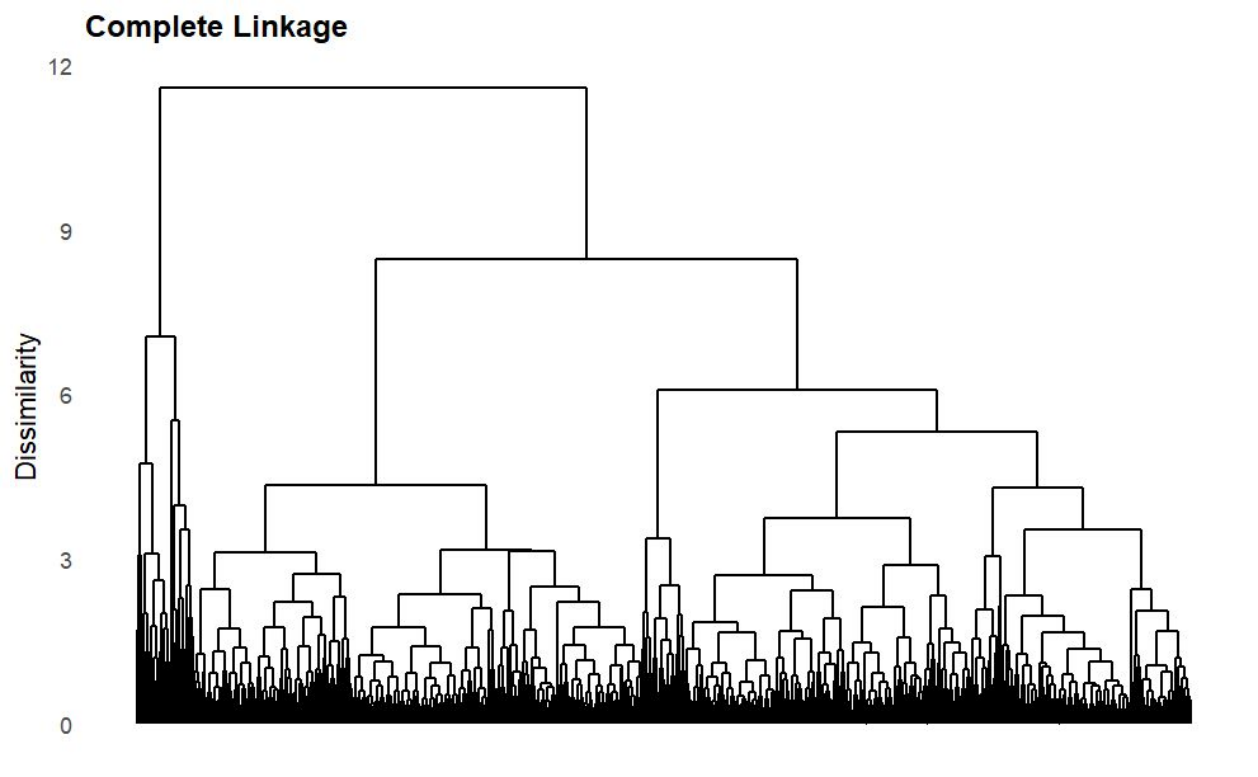
- Baseline models:
  - Linear regression
  - Regularized regression (Ridge, Lasso, Elastic Net)
- Justification: Response variable (percent of adults with obesity) is continuous
- Models we won’t use:
  - Logistic regression → for binary outcomes
  - Poisson regression → for count data

Q/A!

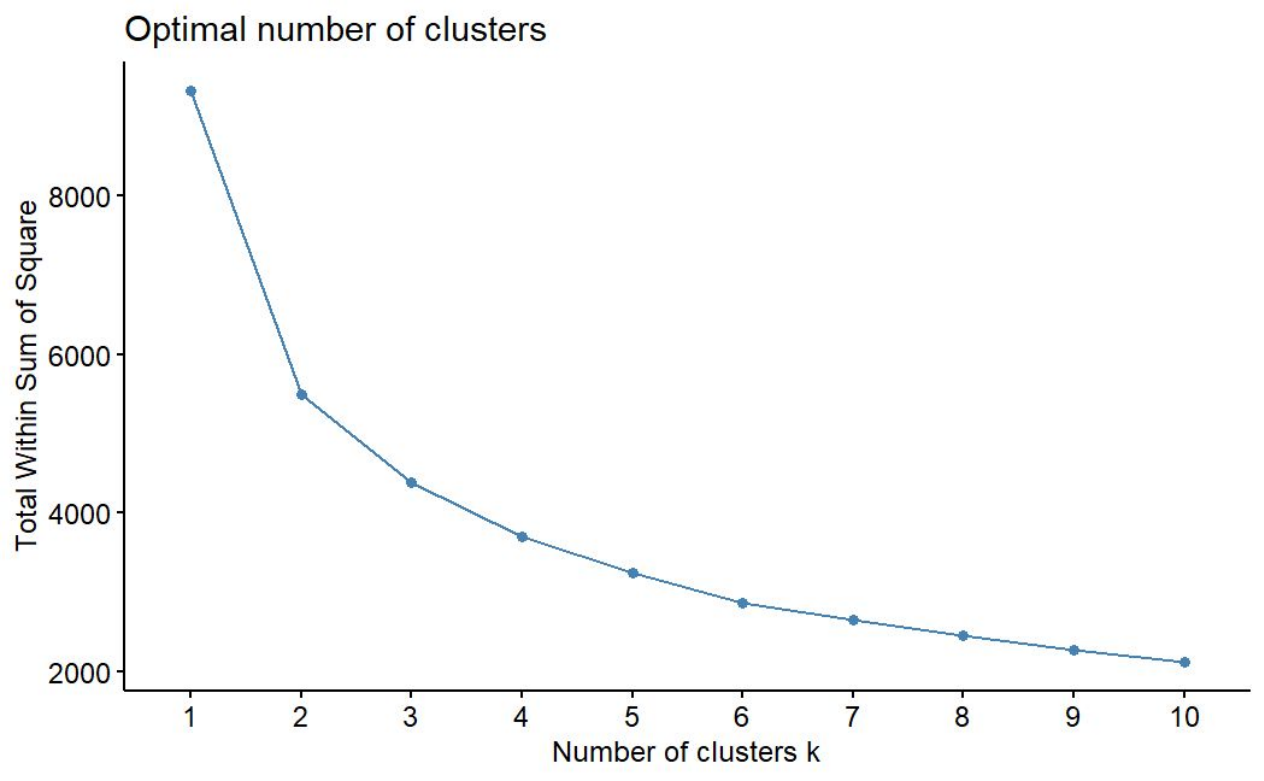
Appendix


Cluster plot
5.0
2.5
0.0
-2.5
Dim2 (19.9%)
-5.0
-2.5
0.0
2.5
5.0
Dim1 (73.3%)
cluster
1
2
3
4

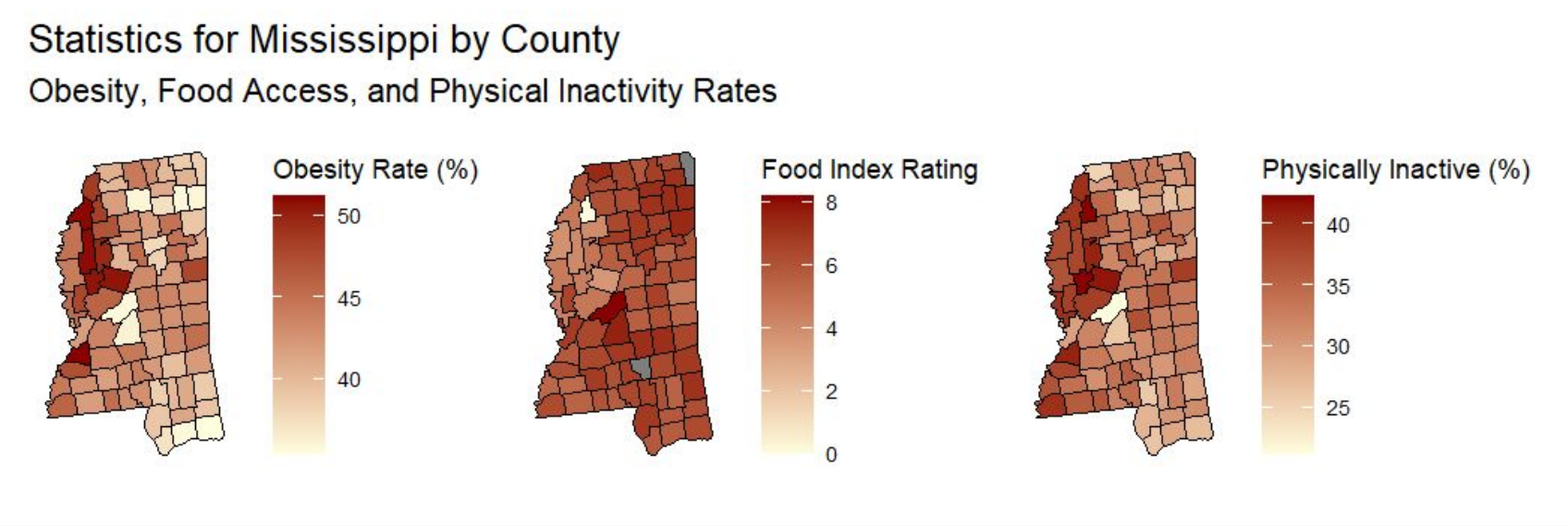
Appendix



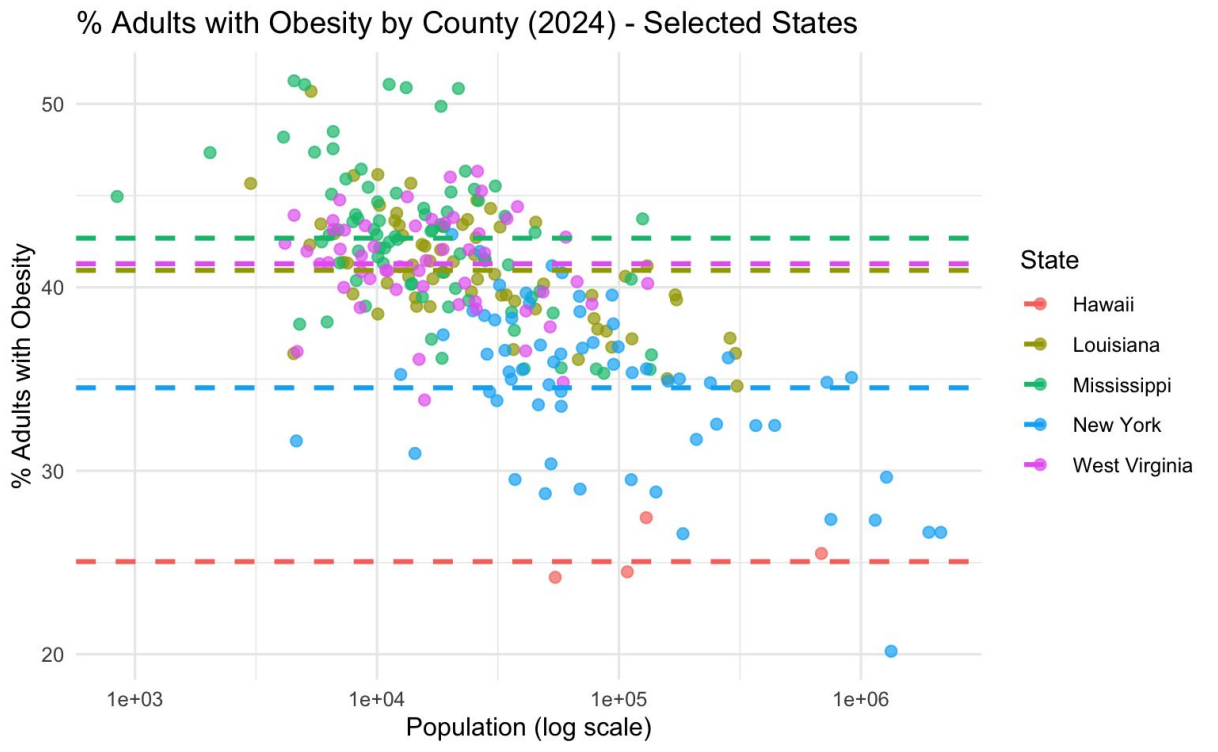
Appendix


Optimal number of clusters
Total Within Sum of Square
8000
6000
4000
2000
1
2
3
4
5
6
7
8
9
10
Number of clusters k

Appendix


Statistics for Mississippi by County
Obesity, Food Access, and Physical Inactivity Rates
Obesity Rate (%)
50
45
40
Food Index Rating
8
6
4
2
0
Physically Inactive (%)
40
35
30
25

Appendix


% Adults with Obesity by County (2024) - Selected States
% Adults with Obesity
50
40
30
20
1e+03
1e+04
1e+05
1e+06
Population (log scale)
State
Hawaii
Louisiana
Mississippi
New York
West Virginia