State-Level Heart Disease Mortality in the United States (Based on Data from 2015-2018)

James Henegan

BDS 726 - Generalized Linear Models

Presentation II

Introduction and Motivation

Motivation

Motivated by Current Events

- As of 4/27, there had been 239 COVID-19 deaths in Mississippi.
- In 137 ($\approx 57\%$) of those cases, the patient had an underlying cardiovascular disease.

General Information

- *Heart disease is the leading cause of death in the United States, causing about 1 in 4 deaths.*[1]
- Here, we examine data related to cardiovascular disease in the United States from 2015 to 2018.

[1] Centers for Disease and Control Prevention

Research Questions

- Is there an association between Heart Disease Mortality and smoking (at the state level)?
- What is the relationship between Heart Disease Mortality, obesity, and median age (at the state level)?
- How do the states “cluster" based on their obesity levels?
- Is there a difference between these two clusters with respect to heart disease mortality?

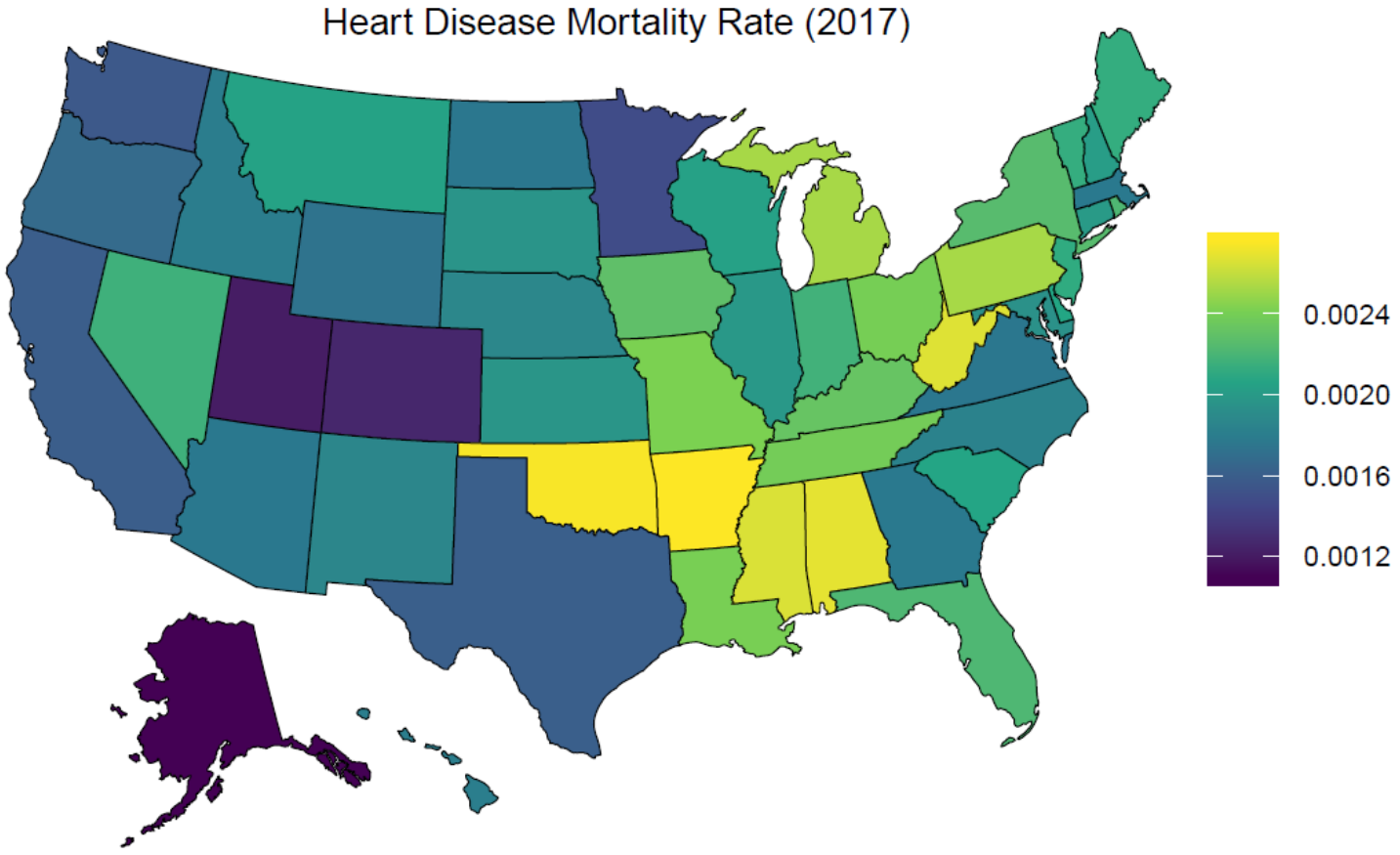
Outcome Measure: Heart Disease Mortality Rate

Example Data - Smoking (2017)

- Let Y_i denote the number of Heart Disease Deaths (as recorded by the Centers for Disease Control and Prevention) for State$_i$ in 2017
- Let Population$_i$ denote the population of State$_i$ in 2017
- Let Smoke$_i$ denote the percentage of adults in State$_i$ with obesity in 2017 (as recorded by the Centers for Disease Control and Prevention)

State$_i$	Y_i	Population$_i$	Smoke$_i$
AL	13110	4874486	20.9
AK	814	739700	21
AZ	12398	7044008	15.6
AR	8270	3001345	22.3
CA	62797	39358497	11.3
CO	7060	5611885	14.6
CT	7138	3573297	12.7
DE	1990	956823	17

Example. We compute the *Heart Disease Mortality Rate* (HDMR) for Alabama (AL) as $13110/4874486 \approx 0.0027$.



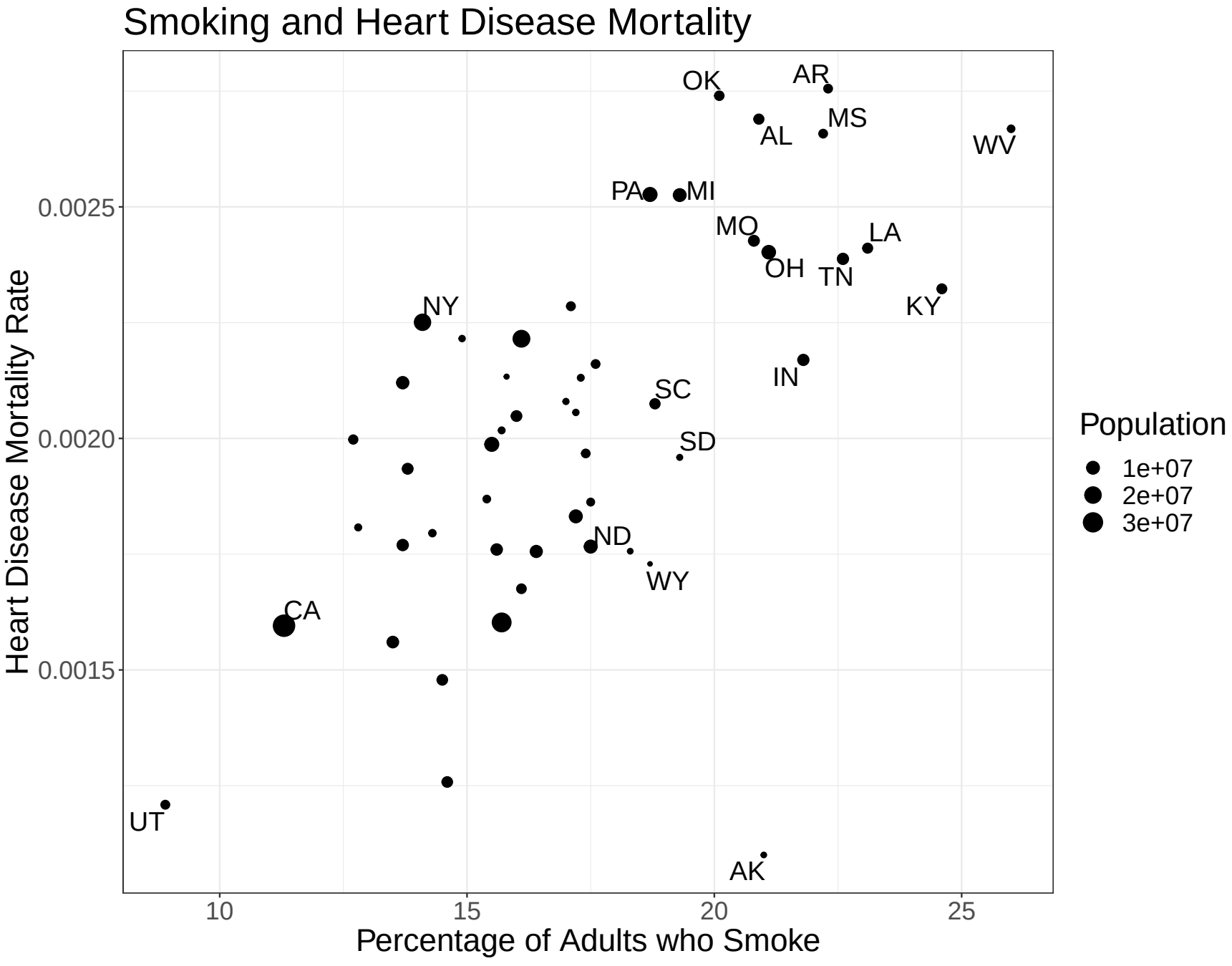
Rank	State	HDMR
1	AR	0.00276
2	OK	0.00274
3	AL	0.00269
4	WV	0.00267
5	MS	0.00266

Rank	State	HDMR
46	WA	0.00156
47	MN	0.00148
48	CO	0.00126
49	UT	0.00121
50	AK	0.00110

Heart Disease Mortality and Smoking in 2017

Research Questions

- Is there an association between Heart Disease Mortality and smoking (at the state level)?
- What is the relationship between Heart Disease Mortality, obesity, and median age (at the state level)?
- How do the states “cluster" based on their obesity levels?
- Is there a difference between these two clusters with respect to heart disease mortality?


Smoking and Heart Disease Mortality
Heart Disease Mortality Rate
0.0025
0.0020
0.0015
10
15
20
25
Percentage of Adults who Smoke
OK
AR
MS
AL
WV
PA
MI
MO
LA
OH
TN
KY
NY
IN
SC
SD
ND
WY
CA
UT
AK
Population
1e+07
2e+07
3e+07

Smoking Model

- Let Y_i denote the number of heart disease deaths for state i
- Let $\texttt{pop}_i$ denote the population of state i
- Let $\texttt{smoke}_i$ denote the percentage of adults who smoke in state i

Suppose that $Y_i \sim \text{Poisson}(\mu_i)$ so that $\text{Var}(Y_i) = \mu_i$.

Consider the model

$$\ln \mu_i = \underbrace{\ln \texttt{pop}_i}_{\text{offset}} + \alpha + \beta \texttt{smoke}_i$$

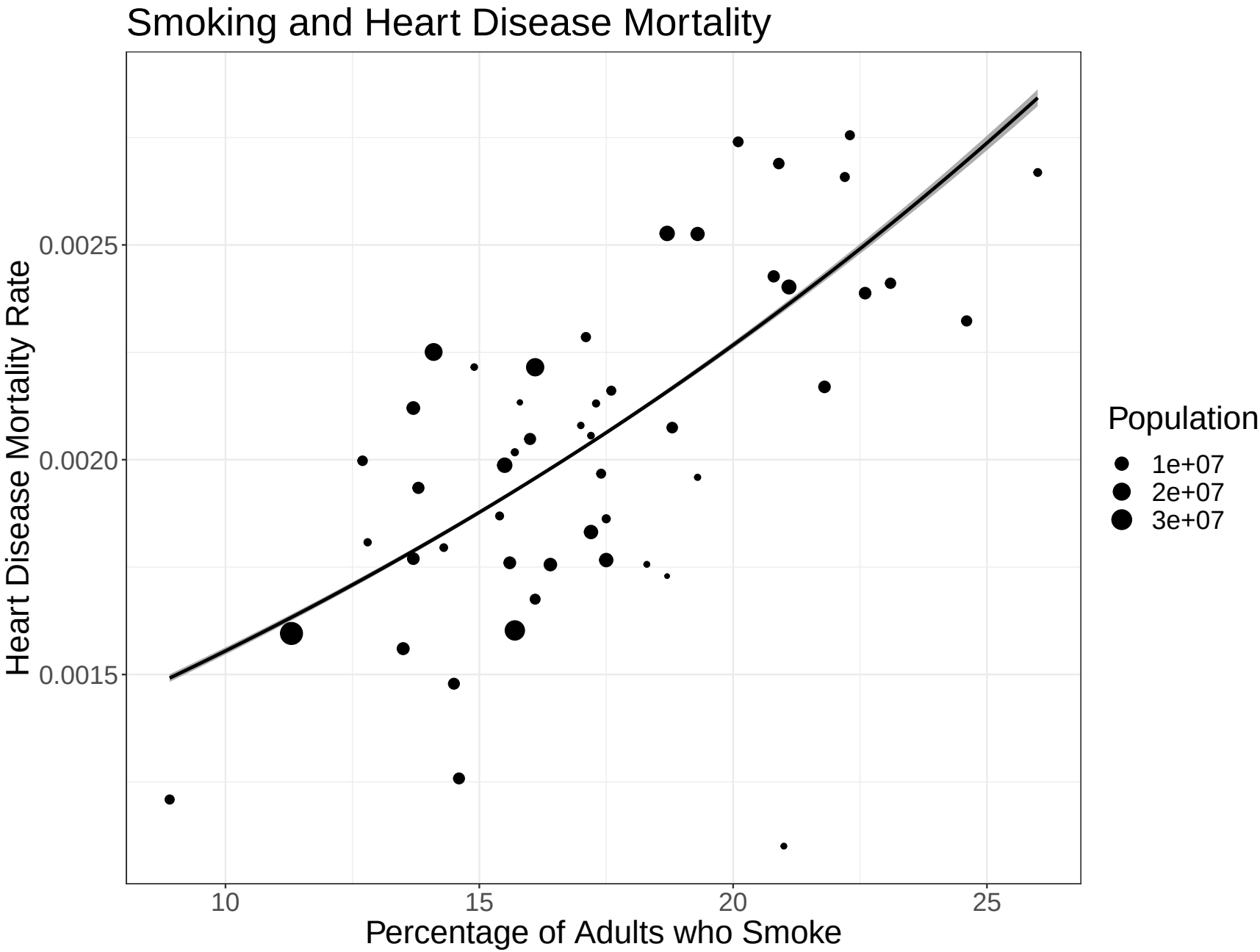
Fit and Interpretation

$$\ln \mu_i = \underbrace{\ln \texttt{pop}_i}_{\text{offset}} + \alpha + \beta \texttt{smoke}_i \quad \Leftrightarrow \quad \frac{\mathrm{E}(Y_i)}{\texttt{pop}_i} = \exp(\alpha + \beta \texttt{smoke}_i)$$

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.8433386  0.0061143 -1119.2   <2e-16
smoke        0.0377018  0.0003562   105.8   <2e-16
```

$$\frac{\mathrm{E}(Y_i | \texttt{smoke}_i = x + 1)}{\mathrm{E}(Y_i | \texttt{smoke}_i = x)} = e^{\hat{\beta}} = e^{0.0377} \approx 1.03842$$

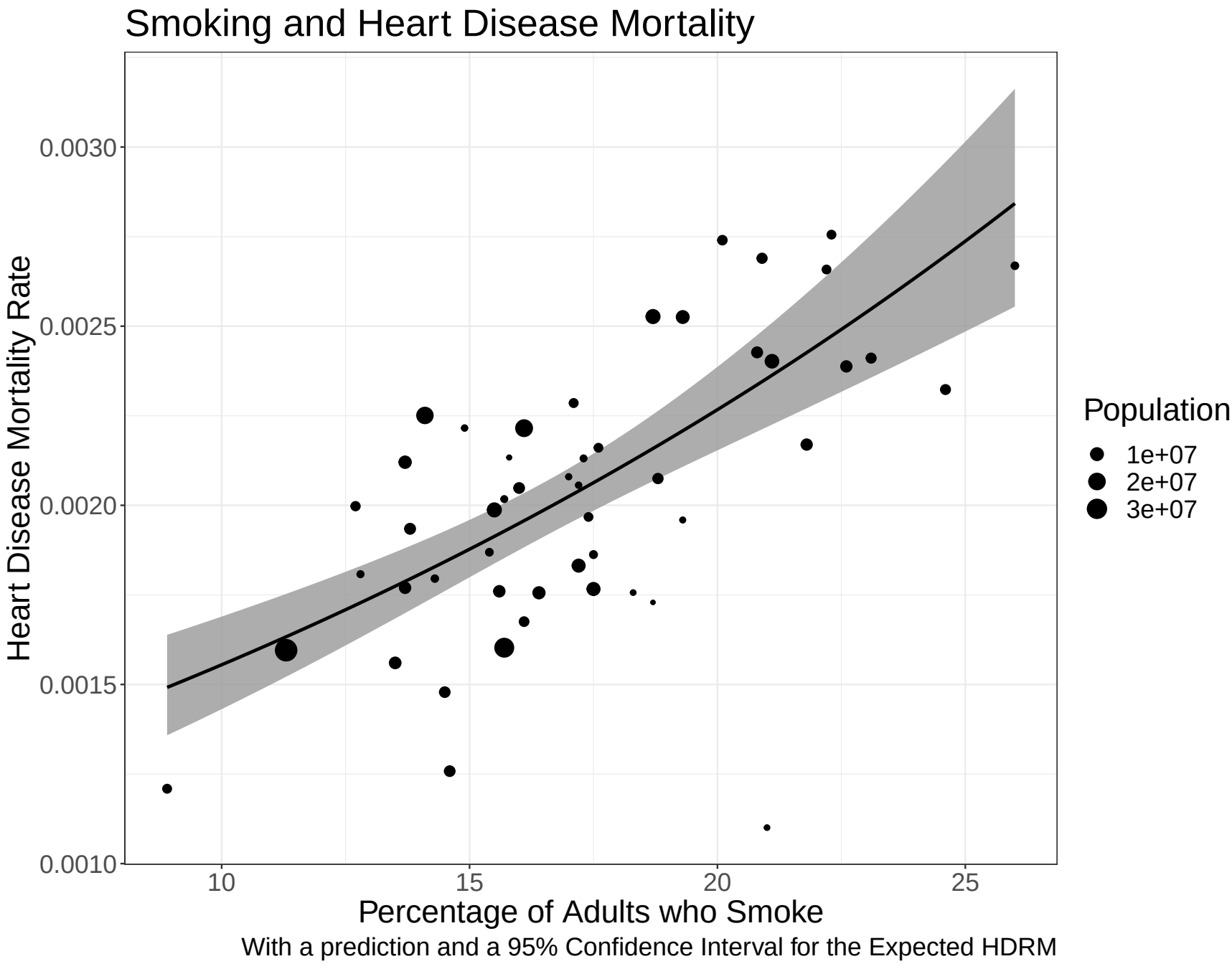
Interpretation: After adjusting for population size, a 1 one unit increase in smoking percentage among adults at the state level is associated with a 4% increase in the number of heart disease deaths

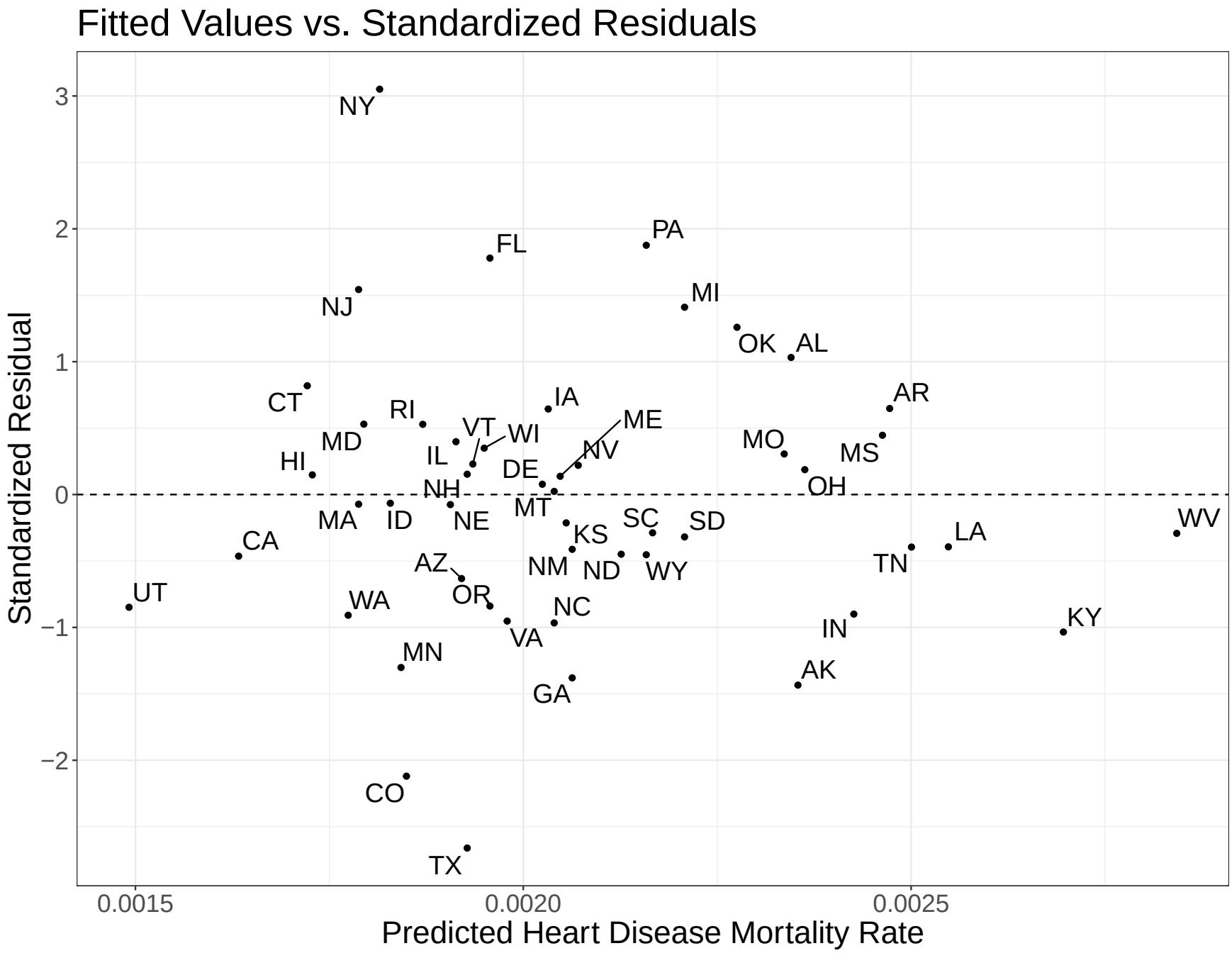

Smoking and Heart Disease Mortality
Heart Disease Mortality Rate
0.0025
0.0020
0.0015
10
15
20
25
Percentage of Adults who Smoke
With a prediction and a 95% Confidence Interval for the Expected HDRM
Population
1e+07
2e+07
3e+07

Overdispersion

- For the Poisson Model, we have Var $Y_i = \mu_i$
- For the so-called "Quasipoisson" Model, we allow Var $Y_i = \phi\mu_i$ for some scaling factor $\phi > 0$
- Fitting the quasipoisson model yields $\hat{\phi} \approx 241.5727$
- The Beta Coefficients for the quasipoisson model are the same as the ones we got for the Poisson model
- The Standard Errors for the coefficients of the quasipoisson model are $\sqrt{241.5727} = 15.5426$ times the Standard Errors for the Poisson model.

	Beta	Poisson S.E.	Quasipoisson S.E.
`(Intercept)`	-6.8433	0.0061	0.0950
`smoke`	0.0377	0.0004	0.0055


Smoking and Heart Disease Mortality
Heart Disease Mortality Rate
0.0030
0.0025
0.0020
0.0015
0.0010
10
15
20
25
Percentage of Adults who Smoke
With a prediction and a 95% Confidence Interval for the Expected HDRM
Population
1e+07
2e+07
3e+07


Fitted Values vs. Standardized Residuals
Standardized Residual
Predicted Heart Disease Mortality Rate
3
2
1
0
−1
−2
0.0015
0.0020
0.0025
NY
FL
PA
NJ
MI
OK
AL
CT
IA
AR
RI
VT
WI
ME
MD
NV
MO
MS
HI
IL
DE
OH
NH
MT
MA
ID
NE
SC
SD
WV
KS
LA
CA
AZ
NM
ND
WY
TN
UT
WA
OR
NC
KY
IN
VA
MN
AK
GA
CO
TX

Negative Binomial Model vs. Quasi-Poisson Model

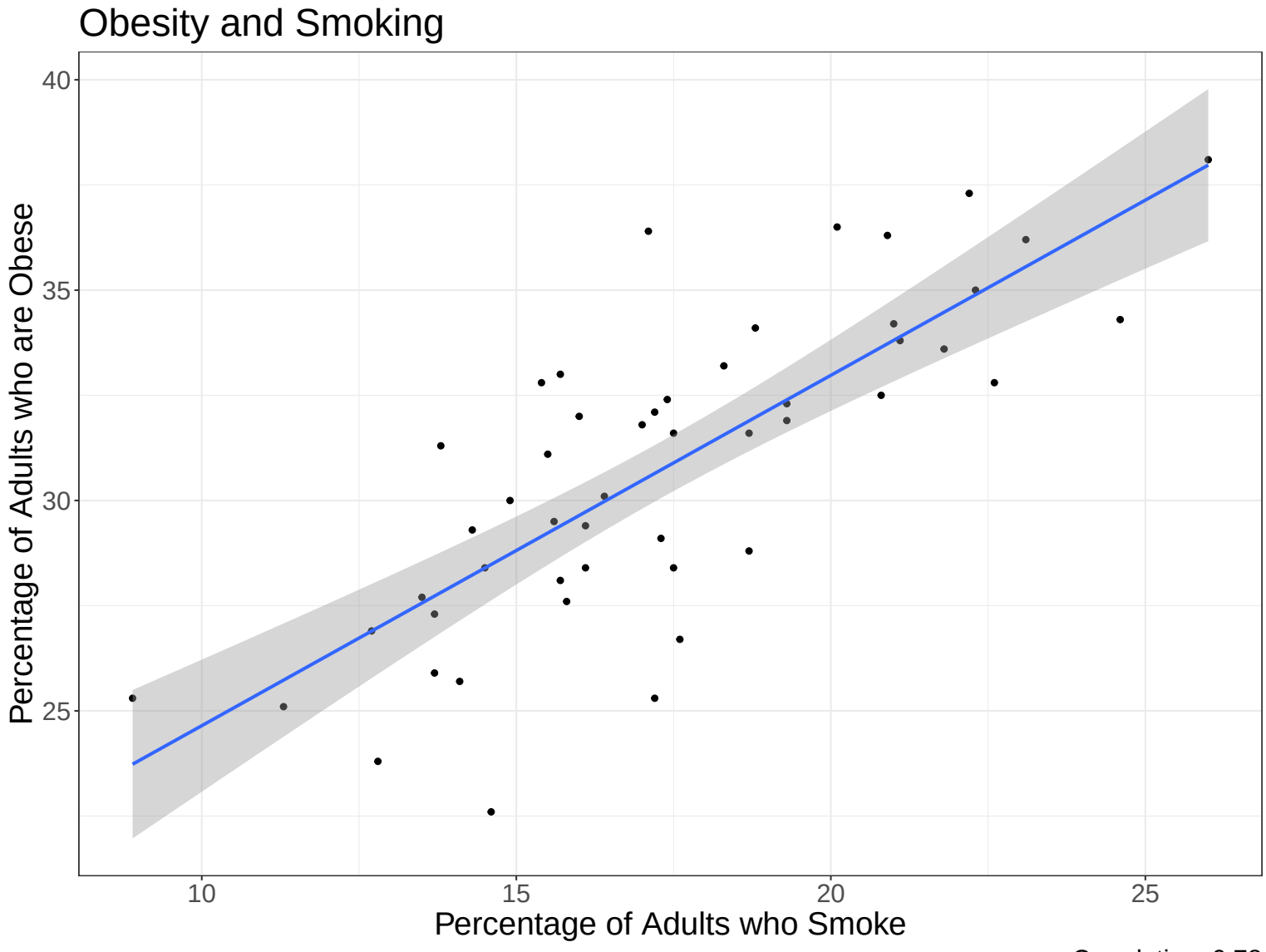
Negative Binomial Model

	Beta	S.E.
(Intercept)	-6.809	0.1127
smoke	0.0348	0.0064

Quasi-Poisson Model

	Beta	S.E.
(Intercept)	-6.8433	0.0950
smoke	0.0377	0.0055

Should we include Obesity in our Smoking Model?



Correlation: 0.78

Comparison of Models: Collinearity and Ill-conditioning

Beta Estimates for Models with Quasipoisson Fits:

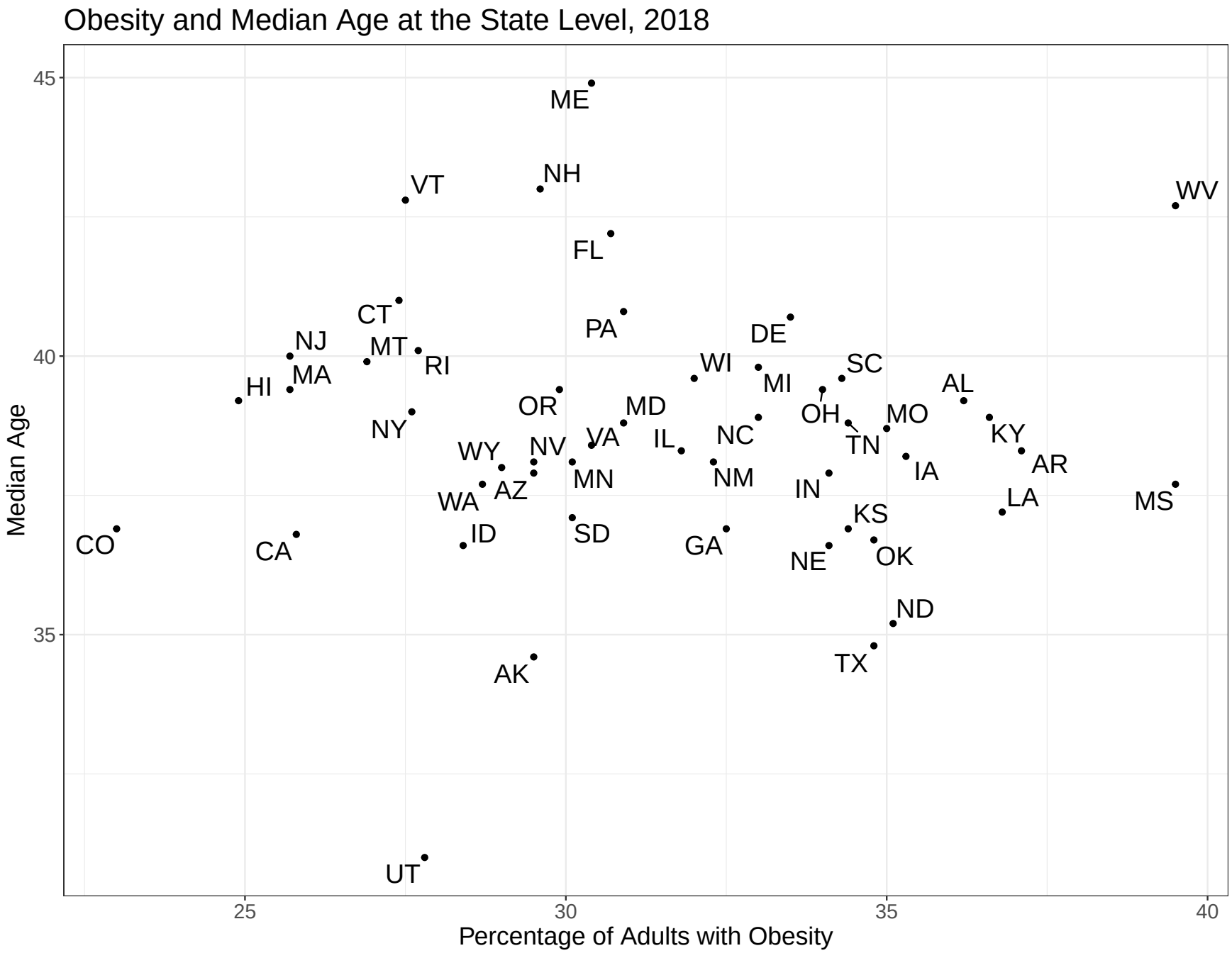
Model	`smoke` estimate	`obesity` estimate
Smoking	0.037702	-
Obesity	-	0.026964
Smoking and Obesity	0.044602	-0.00823

- The p-value for each estimate was < 0.001, except for the Obesity estimate from the "Smoking and Obesity" Model, which had a p-value fo 0.382
- Conclusion: It would be inappropriate to include Obesity in our Smoking model

Heart Disease Mortality, Obesity, and Median Age (2018)

Research Questions

- Is there an association between Heart Disease Mortality and smoking (at the state level)?
- What is the relationship between Heart Disease Mortality, obesity, and median age (at the state level)?
- How do the states “cluster" based on their obesity levels?
- Is there a difference between these two clusters with respect to heart disease mortality?


Obesity and Median Age at the State Level, 2018
Median Age
45
40
35
Percentage of Adults with Obesity
25
30
35
40
ME
VT
NH
WV
FL
CT
PA
DE
NJ
MT
RI
WI
SC
MA
MI
AL
HI
OR
MD
OH
MO
NY
VA
IL
NC
TN
KY
WY
NV
AR
IA
MN
NM
AZ
IN
LA
WA
MS
KS
CO
ID
SD
GA
CA
NE
OK
ND
TX
AK
UT

Obesity and Median Age Model (2018)

- Let Y_i denote the number of heart disease deaths for state i
- Let $\texttt{pop}_i$ denote the population of state i
- Let $\texttt{obese}_i$ denote the percentage of adults who are obese in state i
- Let $\texttt{medAge}_i$ denote the median age of state i

Suppose that $Y_i \sim \text{Poisson}(\mu_i)$ and allow $\text{Var}\, Y_i = \phi\mu_i$ for some $\phi > 0$.

Consider the model

$$\ln \mu_i = \ln \texttt{pop}_i + \alpha + \beta \texttt{obese}_i + \gamma \texttt{medAge}_i$$

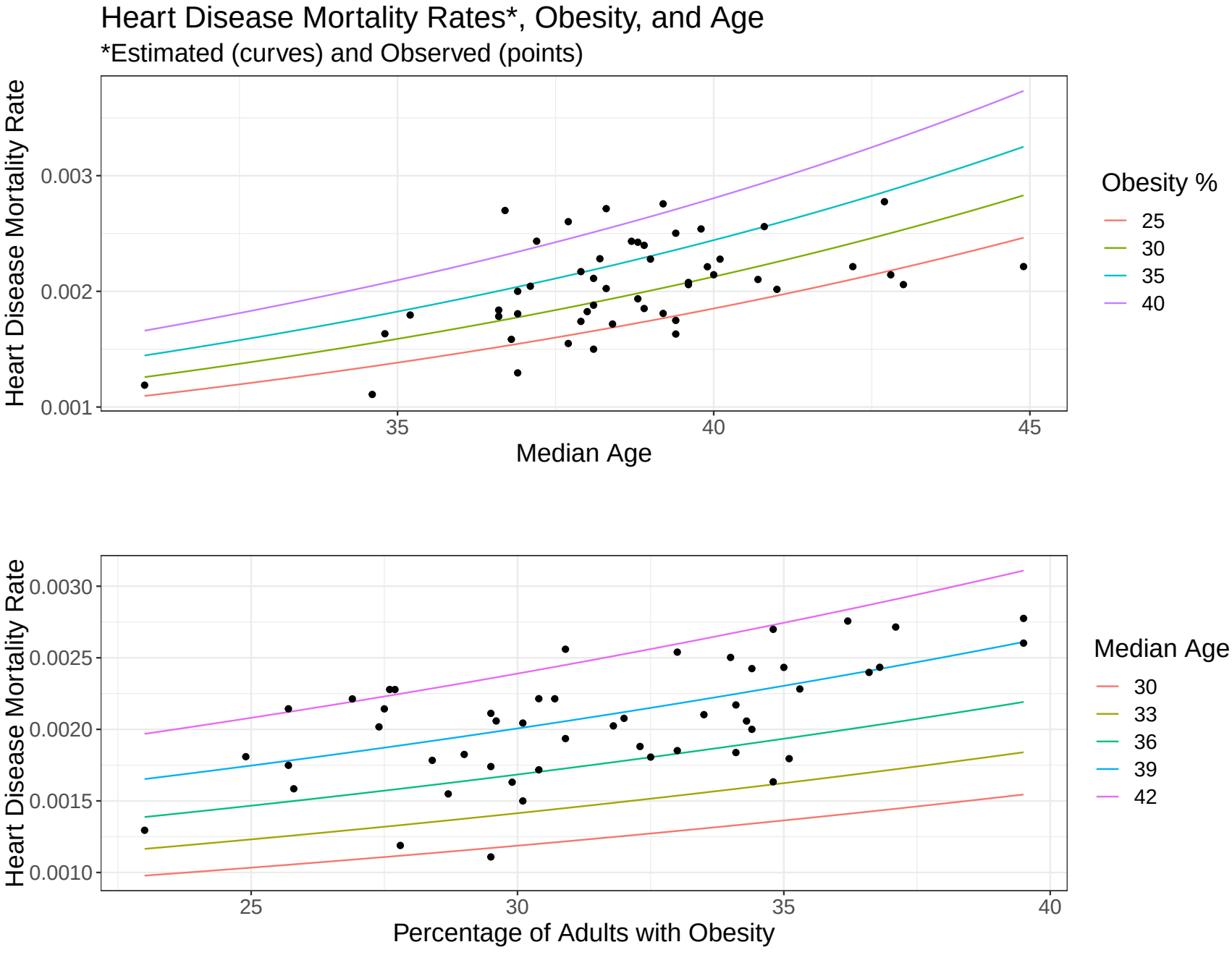
Fit and Interpretation

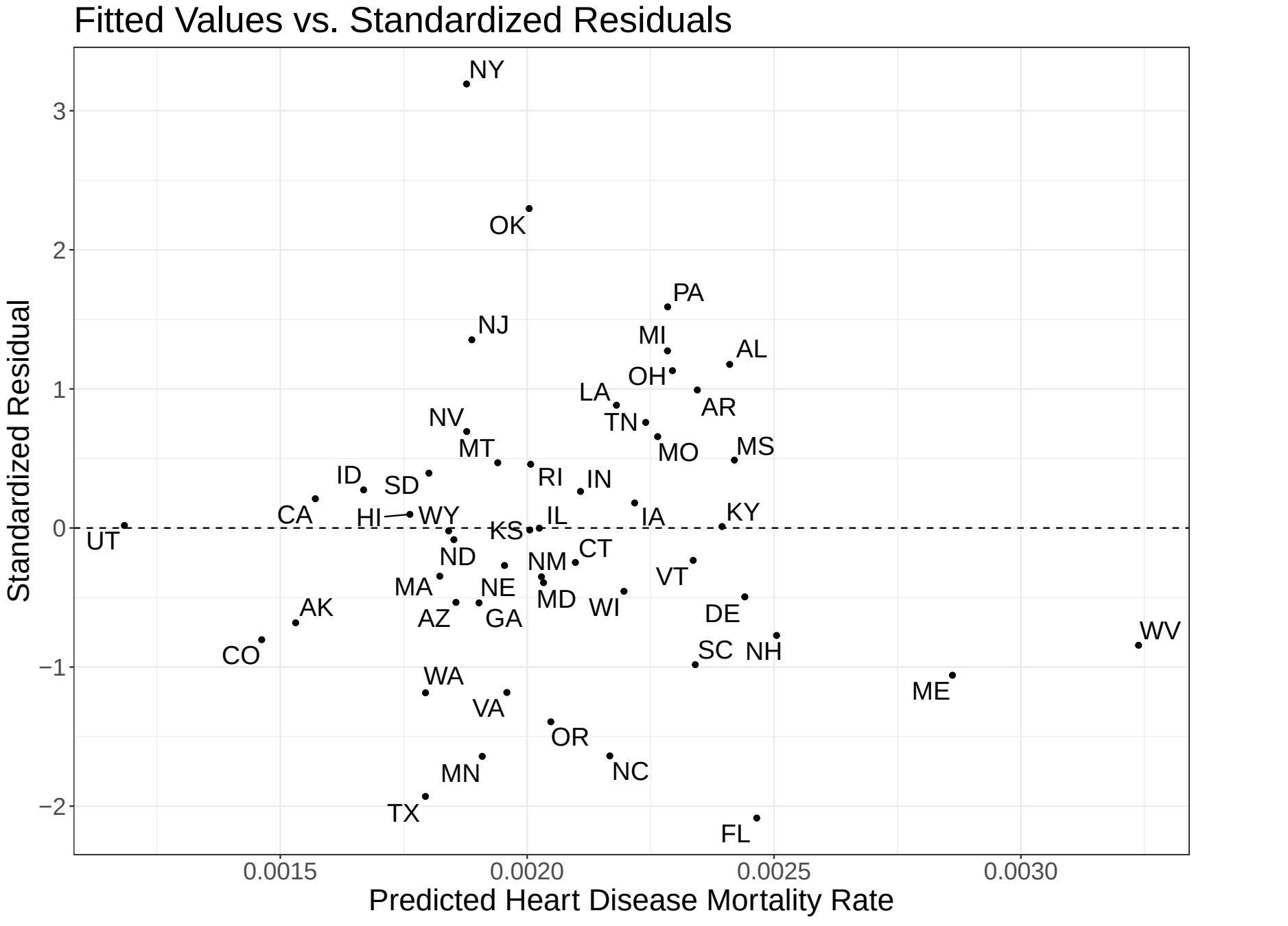
$$\ln \mu_i = \ln \texttt{pop}_i + \alpha + \beta \texttt{obese}_i + \gamma \texttt{medAge}_i$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.315500   0.356440 -26.135  < 2e-16 ***
obese        0.027697   0.004718   5.871 4.22e-07 ***
medAge       0.058285   0.008205   7.103 5.70e-09 ***
```

Dispersion Parameter: $\hat{\phi} \approx 186.0341$

Variable	e^{Estimate}
Obesity Percentage	1.02808
Median Age	1.06002


Heart Disease Mortality Rates*, Obesity, and Age
*Estimated (curves) and Observed (points)
Heart Disease Mortality Rate
0.003
0.002
0.001
35
40
45
Median Age
Obesity %
25
30
35
40
Heart Disease Mortality Rate
0.0030
0.0025
0.0020
0.0015
0.0010
25
30
35
40
Percentage of Adults with Obesity
Median Age
30
33
36
39
42


Fitted Values vs. Standardized Residuals
Standardized Residual
Predicted Heart Disease Mortality Rate
3
2
1
0
−1
−2
0.0015
0.0020
0.0025
0.0030
NY
OK
PA
NJ
MI
AL
OH
LA
AR
NV
TN
MT
MO
MS
ID
SD
RI
IN
CA
HI
WY
IL
IA
KY
UT
KS
ND
CT
NM
VT
MA
NE
MD
WI
AK
AZ
GA
DE
WV
CO
SC
NH
WA
ME
VA
OR
MN
NC
TX
FL

Interaction Term

Let Model 1 denote the model we just considered:

$$\ln \mu_i = \ln \texttt{pop}_i + \alpha + \beta \texttt{obese}_i + \gamma \texttt{medAge}_i$$

Let Model 2 extend Model 1 so that there is an interaction term between age and obesity:

$$\ln \mu_i = \ln \texttt{pop}_i + \alpha + \beta \texttt{obese}_i + \gamma \texttt{medAge}_i + \xi \texttt{obese}_i \times \texttt{medAge}_i$$

Model	DF	Deviance
1	47	8662.7
2	46	8659.7

- H_0: Model 1 fits well compared to Model 2
- Likelihood Ratio Test Statistic: 3.0534 with 1 DF
- p-value ≈ 0.081

Heart Disease and Obesity, 2015-2018

Research Questions

- Is there an association between Heart Disease Mortality and smoking (at the state level)?
- What is the relationship between Heart Disease Mortality, obesity, and median age (at the state level)?
- How do the states “cluster" based on their obesity levels?
- Is there a difference between these two clusters with respect to heart disease mortality?

State-Level Obesity Data

For each state, we have adult obesity percentages for the years 2015-2018.

	state	‘2015‘	‘2016‘	‘2017‘	‘2018‘
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	AL	35.6	35.7	36.3	36.2
2	AK	29.8	31.4	34.2	29.5
3	AZ	28.4	29	29.5	29.5
4	AR	34.5	35.7	35	37.1
5	CA	24.2	25	25.1	25.8
6	CO	20.2	22.3	22.6	23
7	CT	25.3	26	26.9	27.4
8	DE	29.7	30.7	31.8	33.5
9	FL	26.8	27.4	28.4	30.7
10	GA	30.7	31.4	31.6	32.5

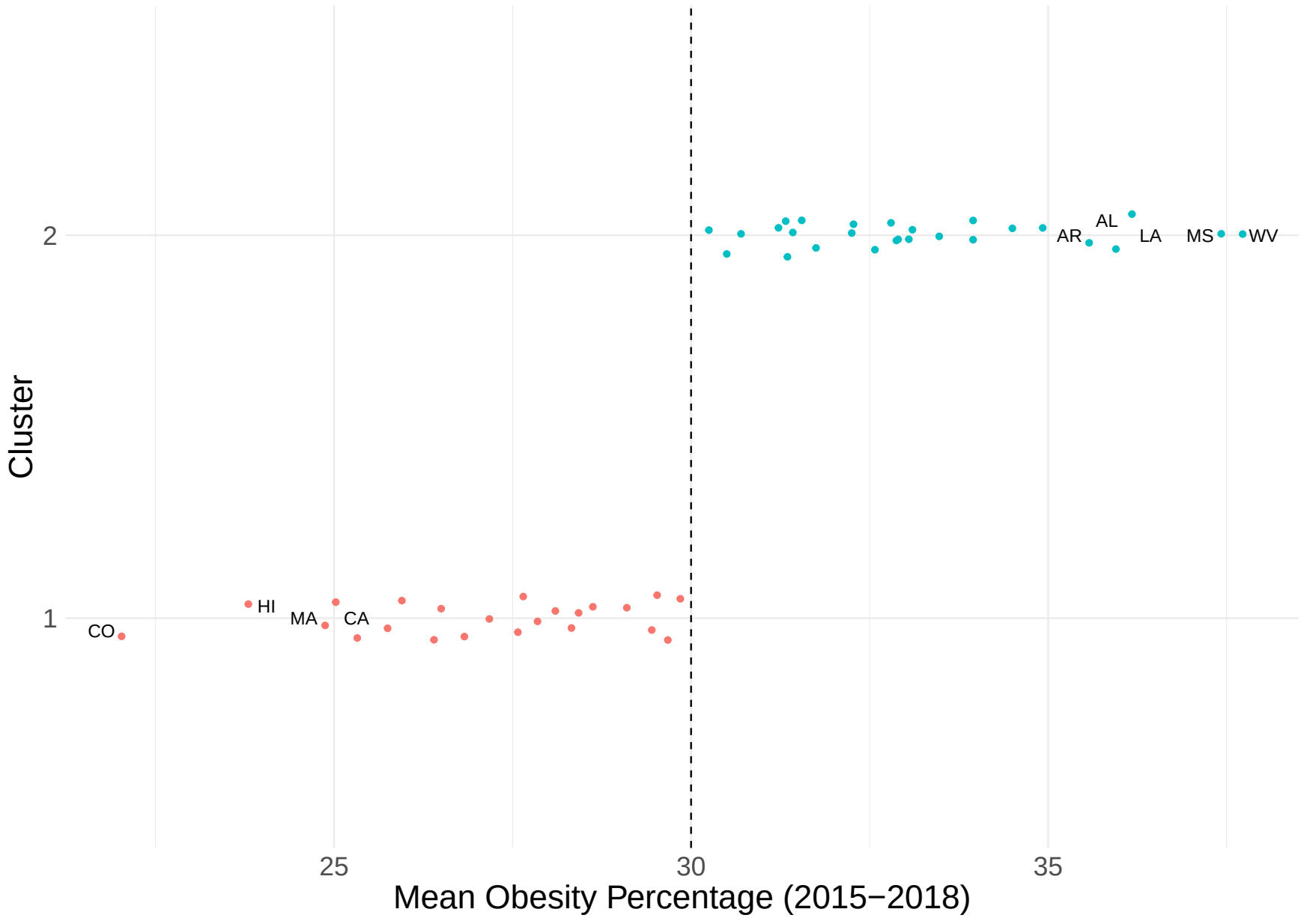
... with 40 more rows

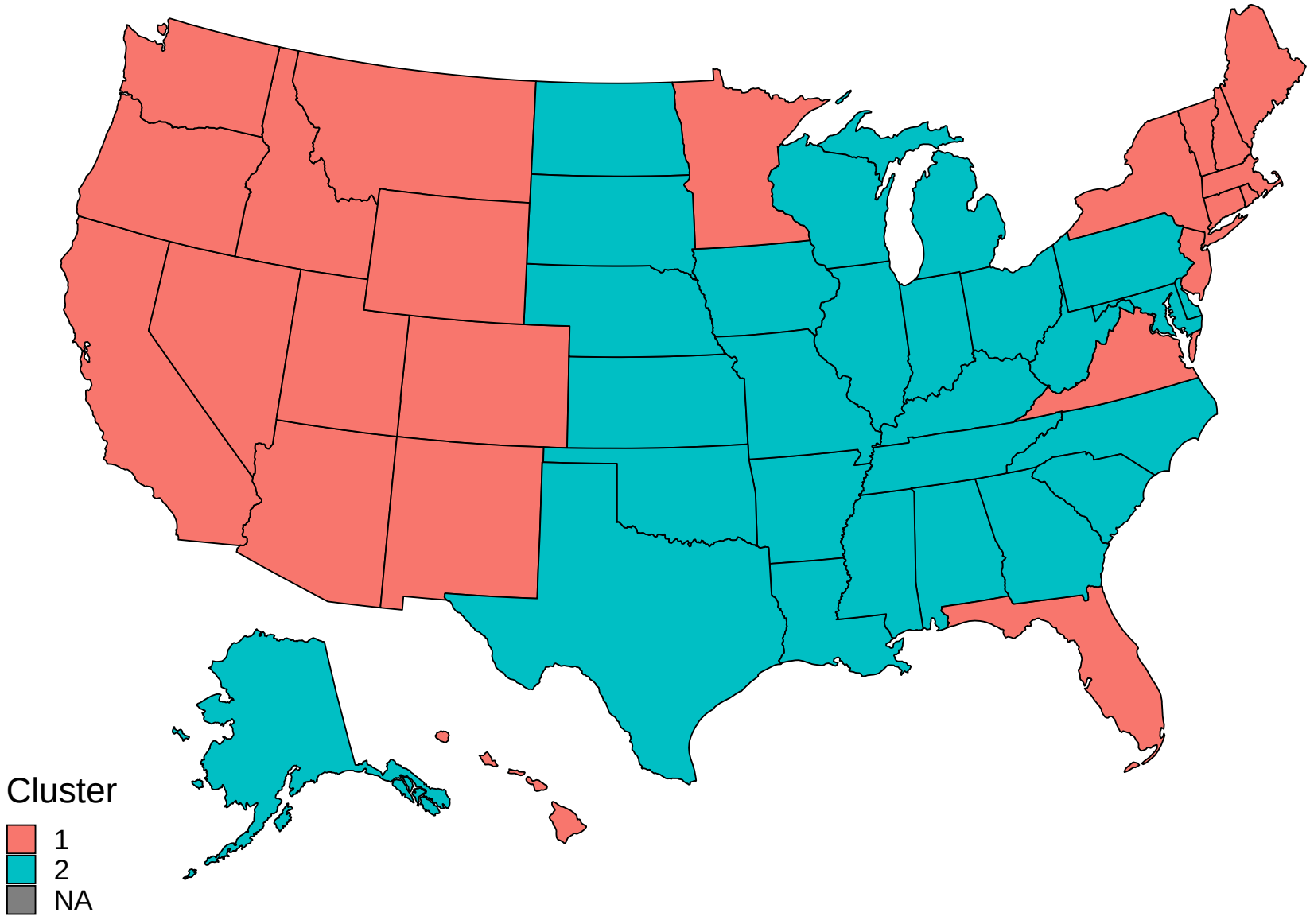
We can use a k-means clustering with $k = 2$ to organize these states into two clusters.

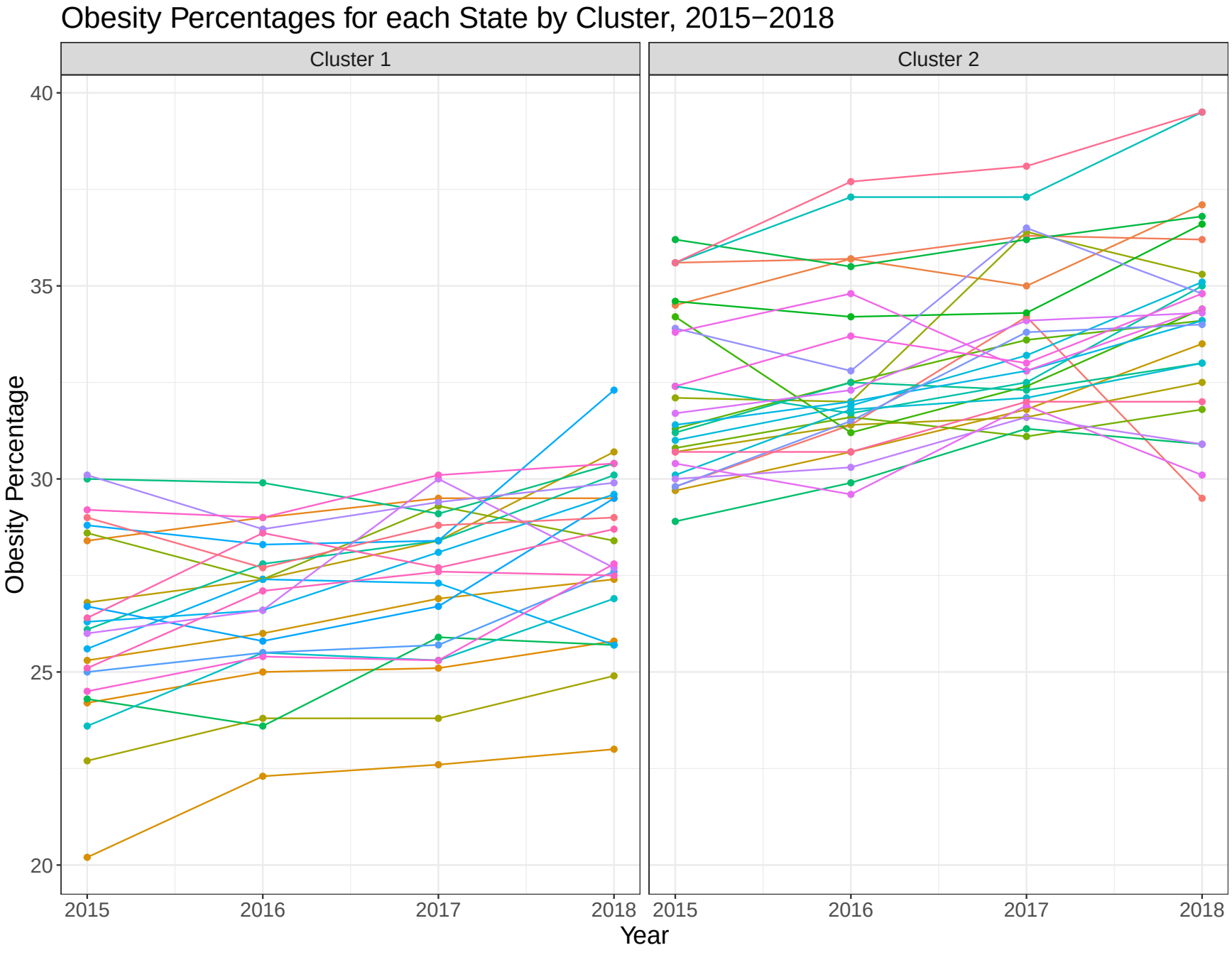
State-Level Obesity Data

```
   state `2015` `2016` `2017` `2018`
   <chr>  <dbl>  <dbl>  <dbl>  <dbl>
 1 AL      35.6   35.7   36.3   36.2
 2 AK      29.8   31.4   34.2   29.5
 3 AZ      28.4   29     29.5   29.5
 4 AR      34.5   35.7   35     37.1
 5 CA      24.2   25     25.1   25.8
# ... with 45 more rows
```

- For these data, it turns out that there is a simple classification rule that will agree with the results of the k-means clustering:
  - Compute the mean obesity percentage for the years 2015-2018 and check to see whether it is less than 30. (Yes $\rightarrow$ Cluster 1.)
- Example: Arizona (AZ) is in Cluster 1. Alabama (AL) is in Cluster 2.

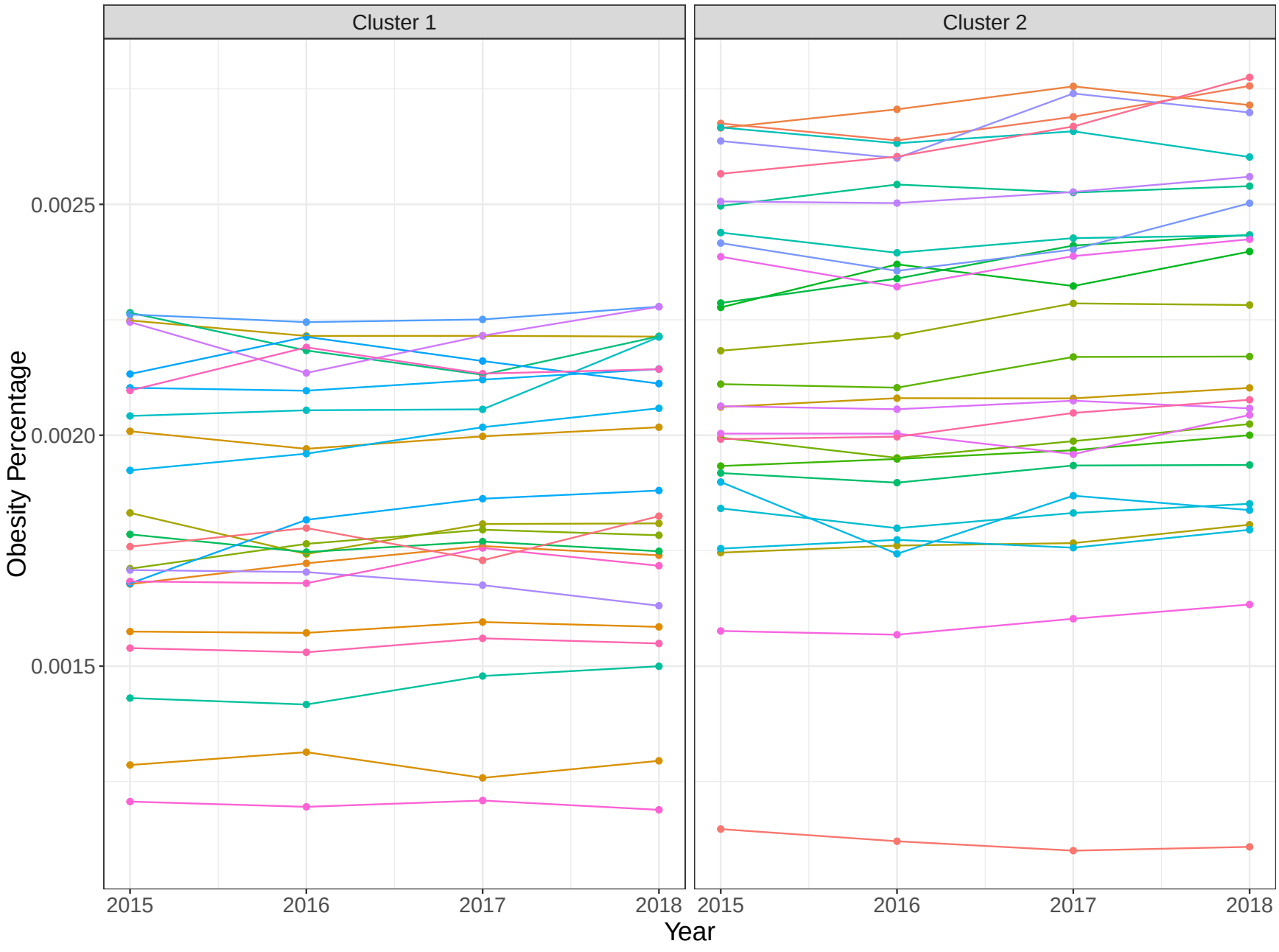

A Simple Classification Rule
that Agrees with our K−Means Clustering
Cluster
2
1
AL
AR
LA
MS
WV
HI
MA
CA
CO
25
30
35
Mean Obesity Percentage (2015−2018)


Cluster
1
2
NA


Obesity Percentages for each State by Cluster, 2015−2018
Cluster 1
Cluster 2
Obesity Percentage
40
35
30
25
20
2015
2016
2017
2018
Year

Research Questions

- Is there an association between Heart Disease Mortality and smoking (at the state level)?
- What is the relationship between Heart Disease Mortality, obesity, and median age (at the state level)?
- How do the states “cluster" based on their obesity levels?
- Is there a difference between these two clusters with respect to heart disease mortality?


Heart Disease Mortality Rates for each State by Cluster, 2015−2018
Cluster 1
Cluster 2
Obesity Percentage
0.0025
0.0020
0.0015
2015
2016
2017
2018
Year

A Repeated Measurement Model

- Each state has four measurements
  - one measurement per year for the years 2015-2018
- Let Y_{ij} denote the number of heart disease deaths for state i at measurement j
- Let $\texttt{pop}_{ij}$ denote the population of state i at measurement j
- Let $\texttt{obese}_{ij}$ denote the percentage of adults who are obese in state i at measurement j
- Suppose that $Y_{ij} \sim \text{Poisson}(\mu_{ij})$
- $\text{Var}(Y_{ij}) = \phi\mu_{ij}$ for some $\phi > 0$
- $\text{Corr}(Y_{ij}, Y_{ik}) = \alpha$

Consider the model

$$\ln \mu_{ij} = \ln \texttt{pop}_{ij} + \alpha + \beta \texttt{cluster2}_{ij}$$

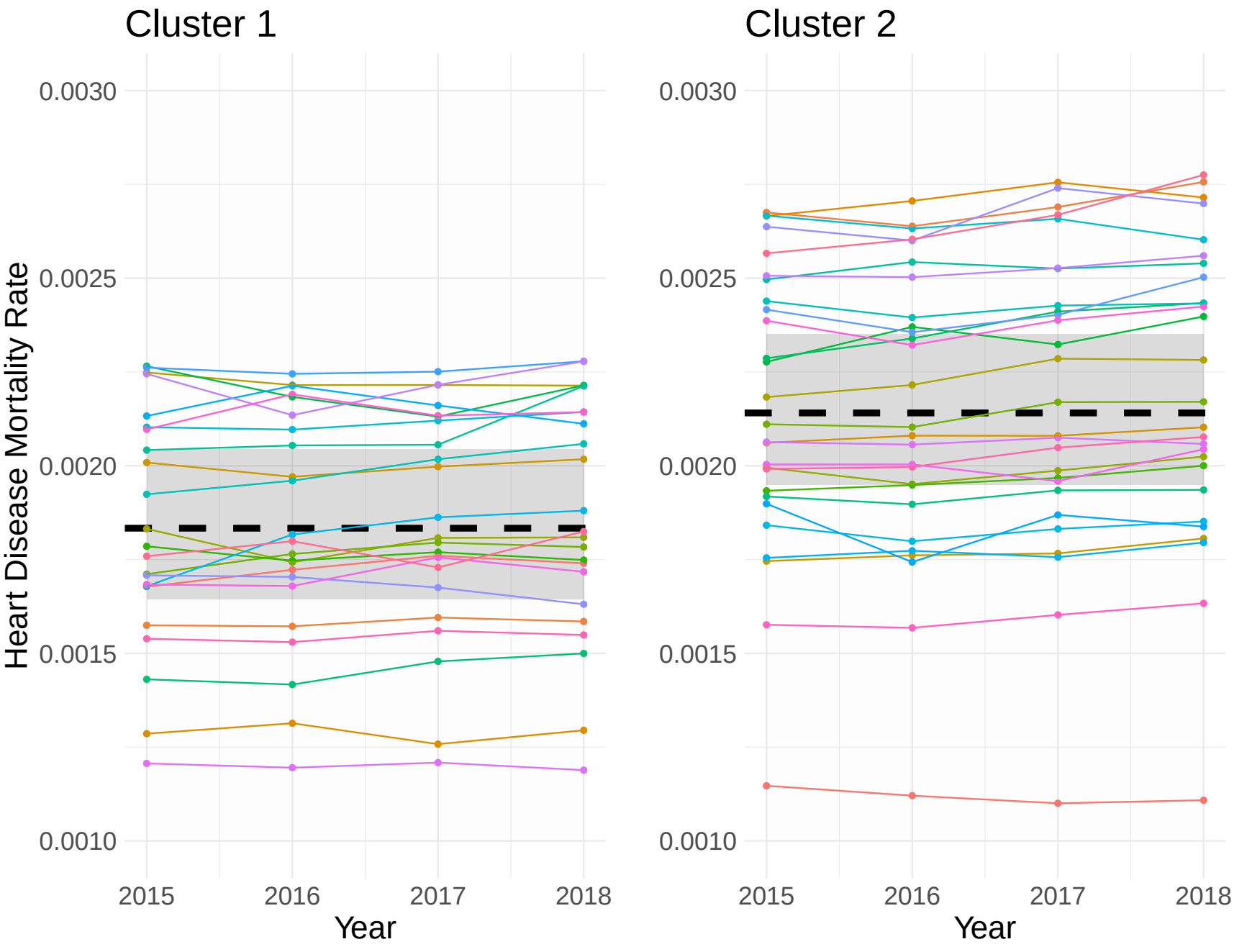
Results of Fitting the Model (Empirical SE Estimates)

$$\ln \mu_{ij} = \ln \texttt{pop}_{ij} + \alpha + \beta \texttt{cluster2}_{ij}$$

	Estimate	SE	95% LB	95% UB	p-value
α	-6.3015	0.0555	-6.4102	-6.1928	$< .0001$
β	0.1549	0.0732	0.0114	0.2985	0.0344

- $\exp(-6.3015) \approx 0.0018$
- $\exp(-6.3015 + 0.1549) \approx 0.0021$
- $\exp(0.1549) \approx 1.17$

Based on this model, the expected heart disease mortality rate for a state in obesity-cluster 2 is approximately 17% greater than the the expected heart disease mortality rate for a state in obesity-cluster 1.


Cluster 1
Cluster 2
Heart Disease Mortality Rate
0.0030
0.0025
0.0020
0.0015
0.0010
2015
2016
2017
2018
Year

Closing Comments

Closing Comments

- In epidemiology, an *ecological fallacy* can occur when we make an inference about an individual while using data that has been aggregated at the group level.
- In this presentation, we have examined the relationship between obesity and heart disease mortality at the state level, not the individual level.
- More detailed research is required to understand the relationship between smoking, obesity, and heart disease at the individual level.
- Nevertheless, it is still a good idea to “play it safe" and

take care of your heart!

Thank You!