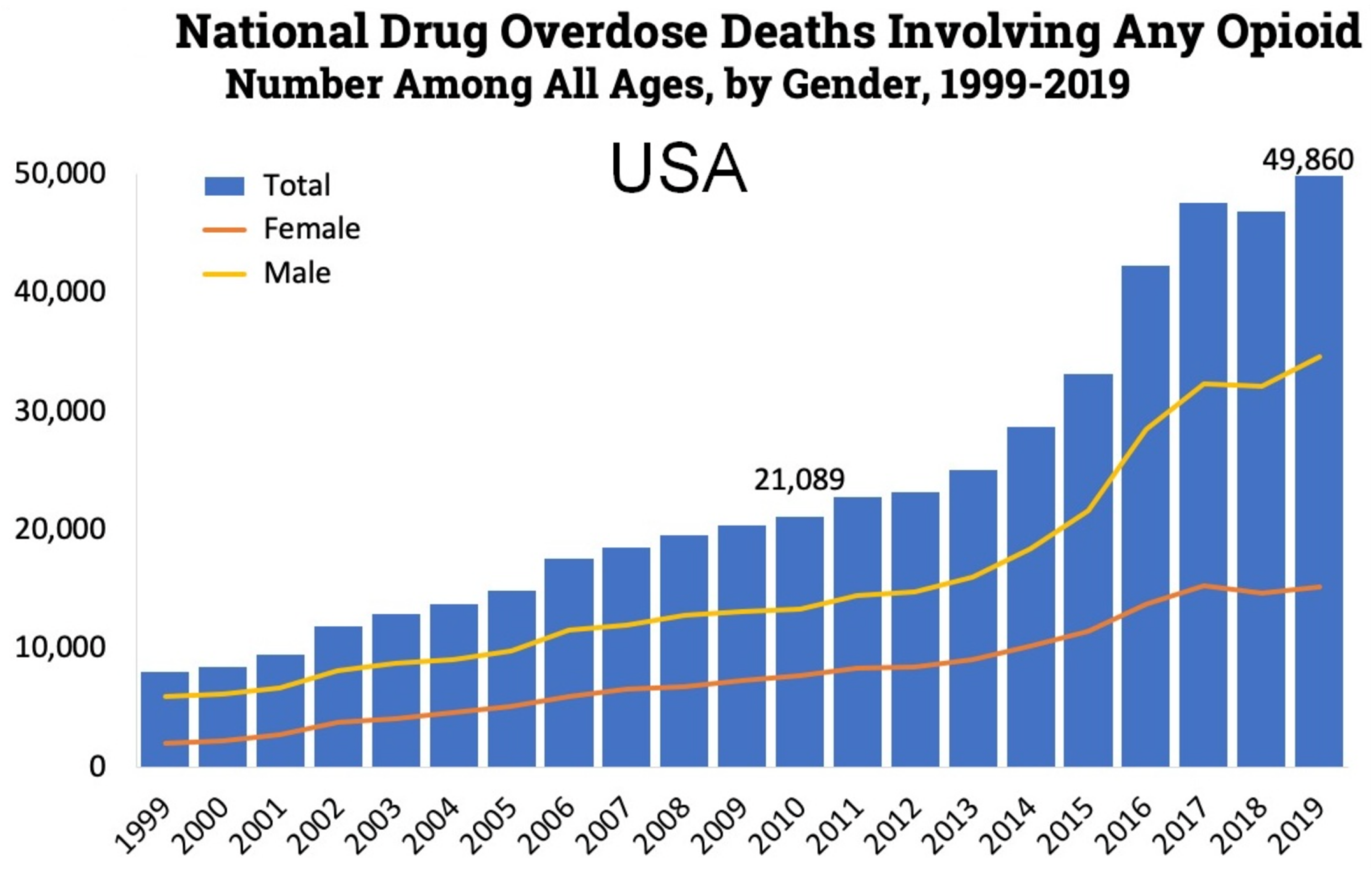
Modeling the Opioid Epidemic in America

Stone Wang

Background

- Moderate to strong painkillers including oxycodone, hydrocodone, and fentanyl
- More than 70,000 Americans died from any/all drug-involved overdose in 2019
  - 21,088 opioid deaths in 2010, 49,860 opioid deaths in 2019
- Rural PA and Philadelphia suffer from highest rate of drug overdose in the country



https://www.drugabuse.gov/drug-topics/trends-statistics/overdose-death-rates

Goal

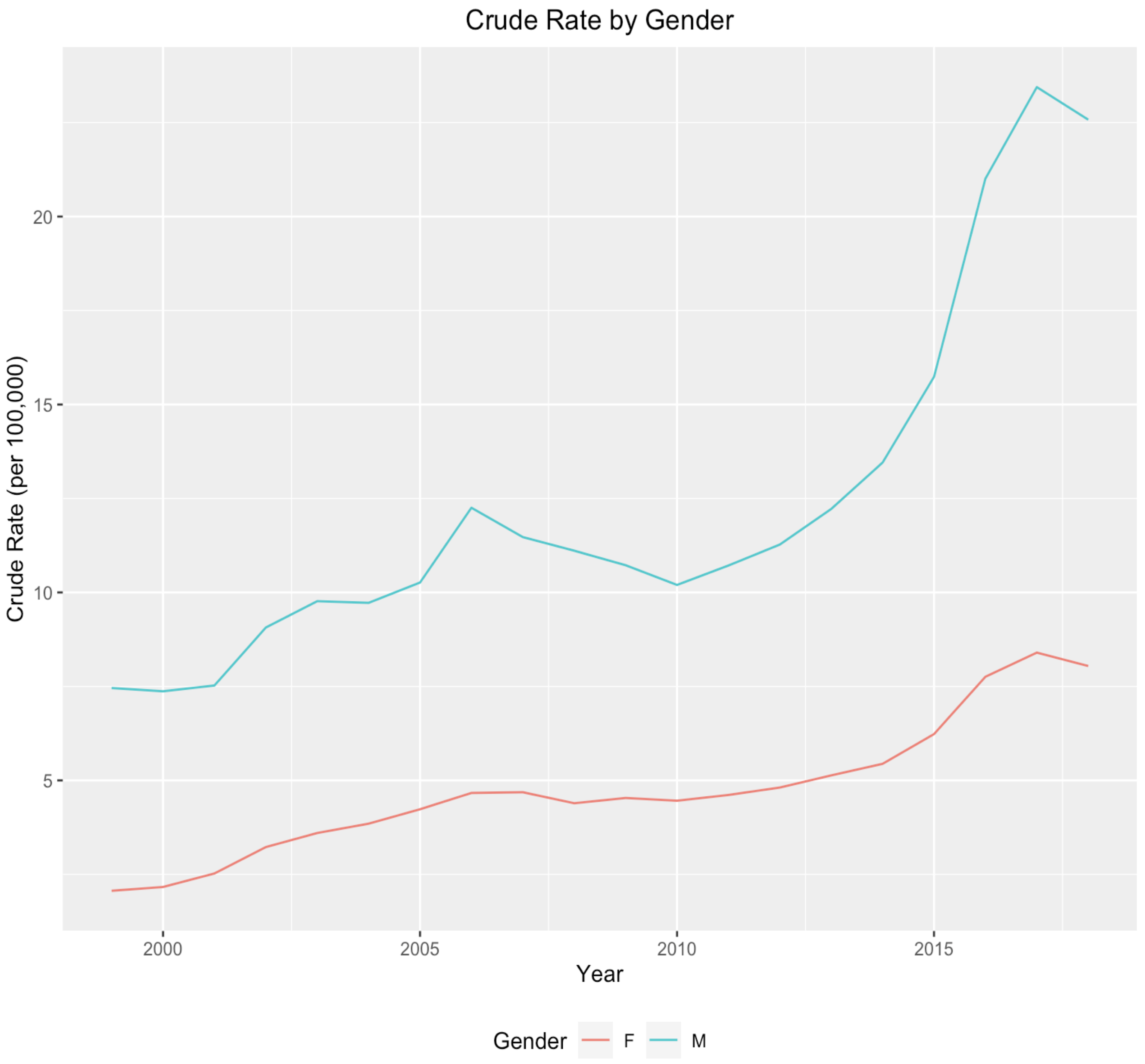
- Discover the demographics impacted the most
- Compare Bayesian and frequentist approaches
- Explore variable selection methods with interaction terms

CDC Wonder Data

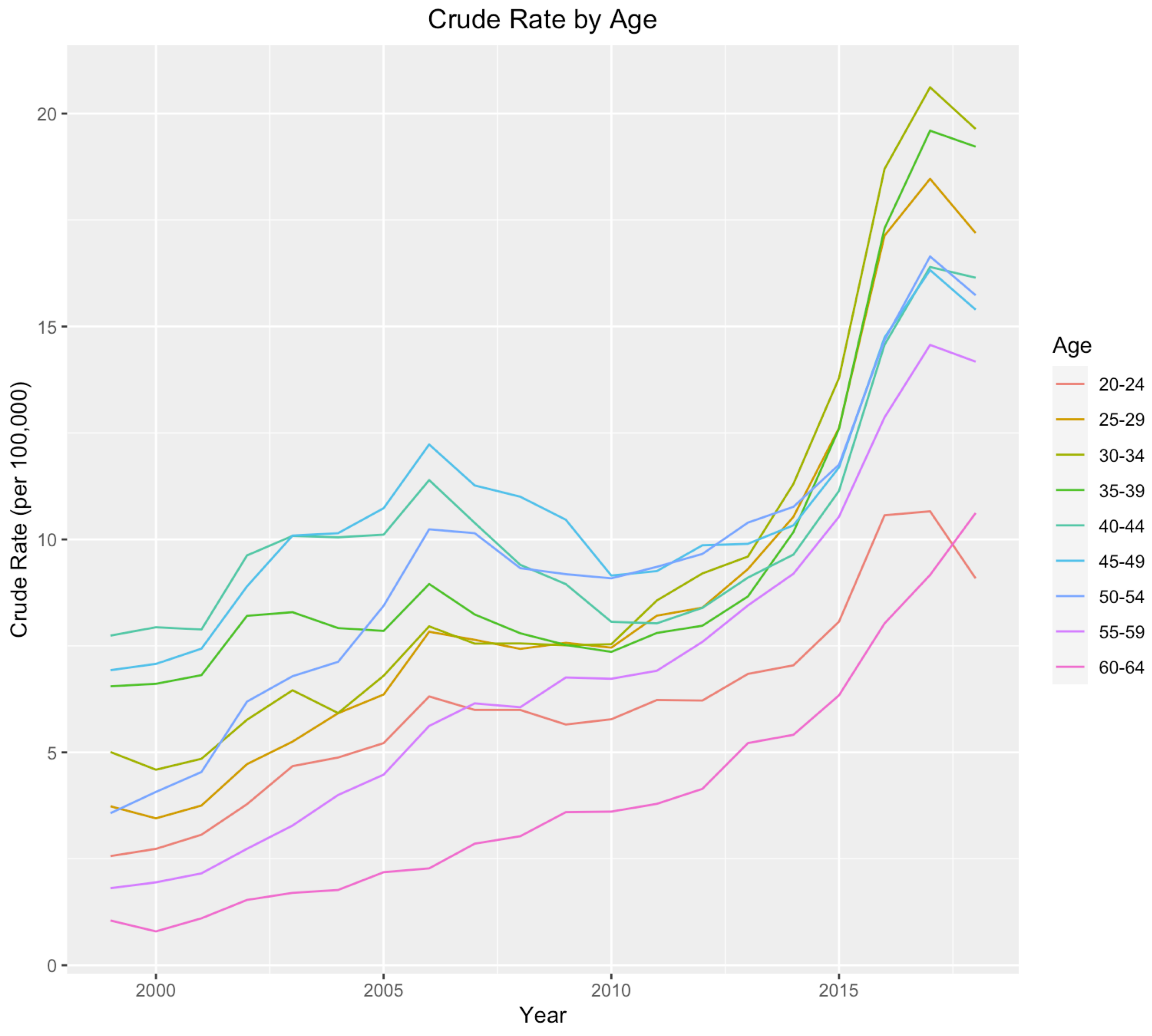
- Online database with ad-hoc query system for analysis of public health data
- Data range from 1999 to 2018
- Can only export 5 factors:
  - Gender: Male, Female
  - Race: American Indian/Alaska Native, Asian, African, White
  - Age group: 5-year groups from 20 to 64
  - Deaths
  - Population
- Response: crude rate (mortality rate) = Deaths/Population

Year	Gender	Race	Age	Deaths	Population	Crude Rate	CR.100
2005	M	White	55-59	369	7233416	5.10E-05	5.10132
2005	F	White	60-64	74	5812425	1.27E-05	1.27313
2005	M	White	60-64	136	5418172	2.51E-05	2.51007
2006	F	American Indian or Alaska Native	20-24	1	154010	6.49E-06	0.64931
2006	M	American Indian or Alaska Native	20-24	8	171339	4.67E-05	4.66911

EDA

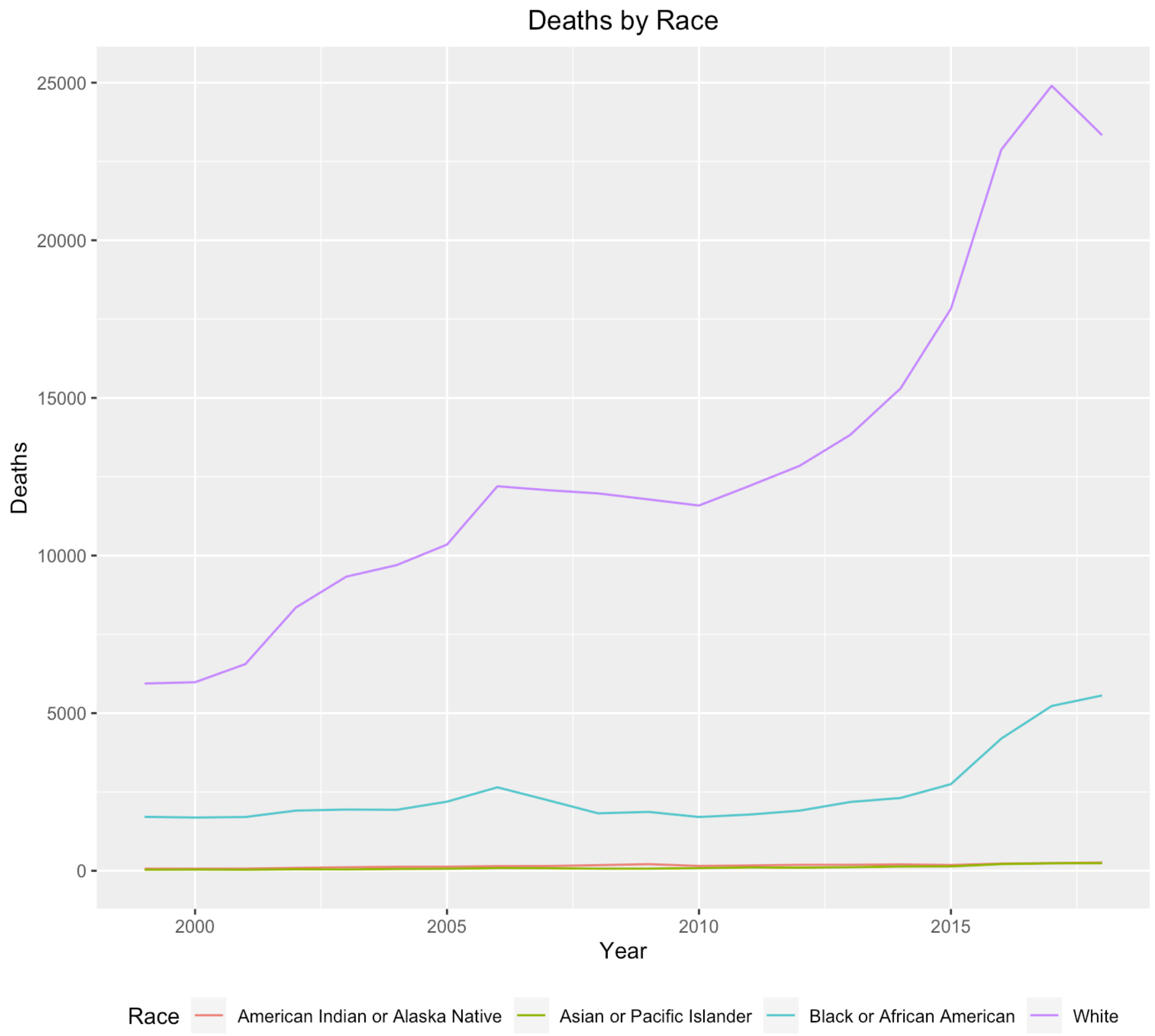


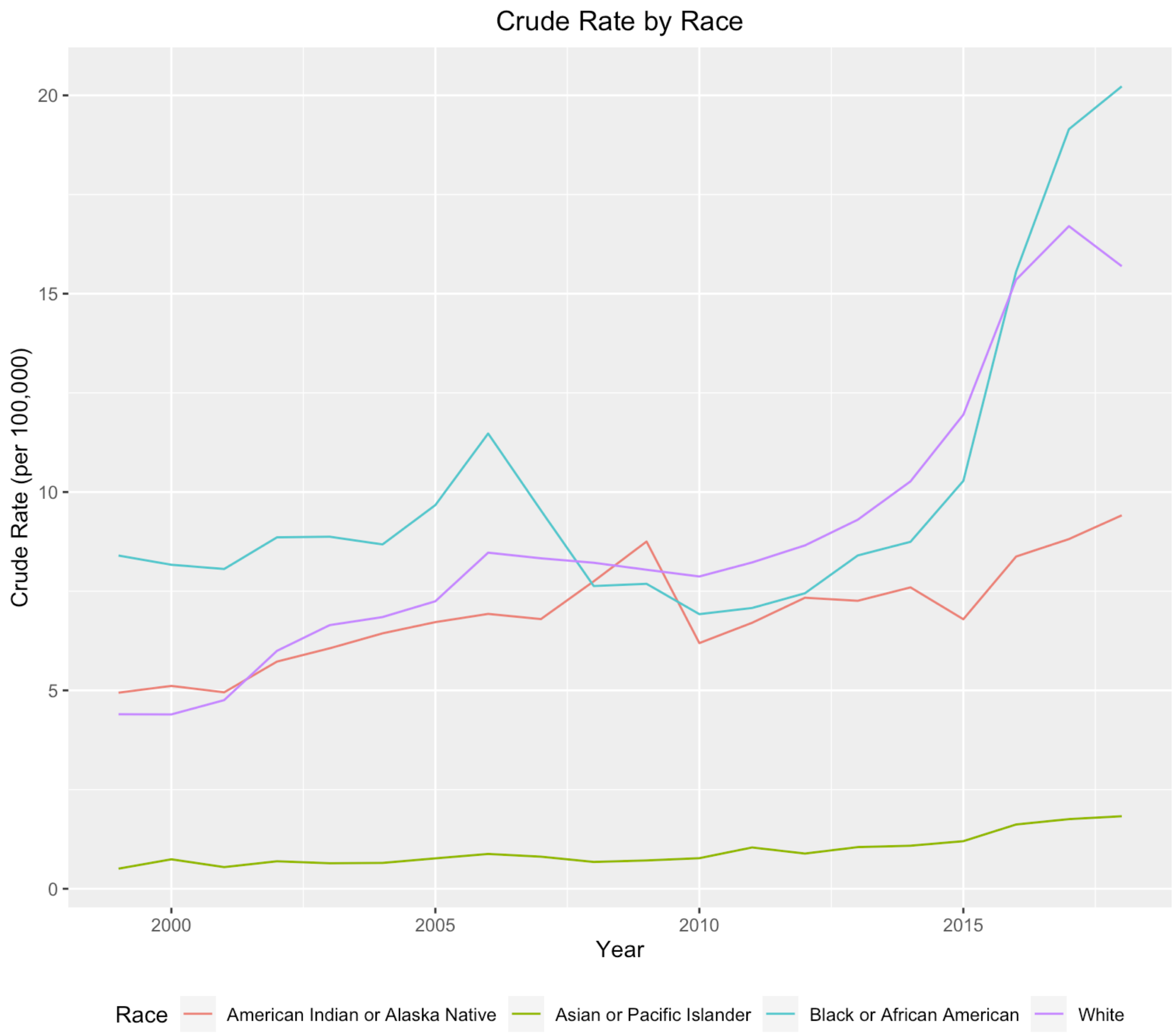
Crude Rate by Gender
Crude Rate (per 100,000)
20
15
10
5
2000
2005
2010
2015
Year
Gender
F
M



Crude Rate by Age
Crude Rate (per 100,000)
20
15
10
5
0
2000
2005
2010
2015
Year
Age
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64

EDA

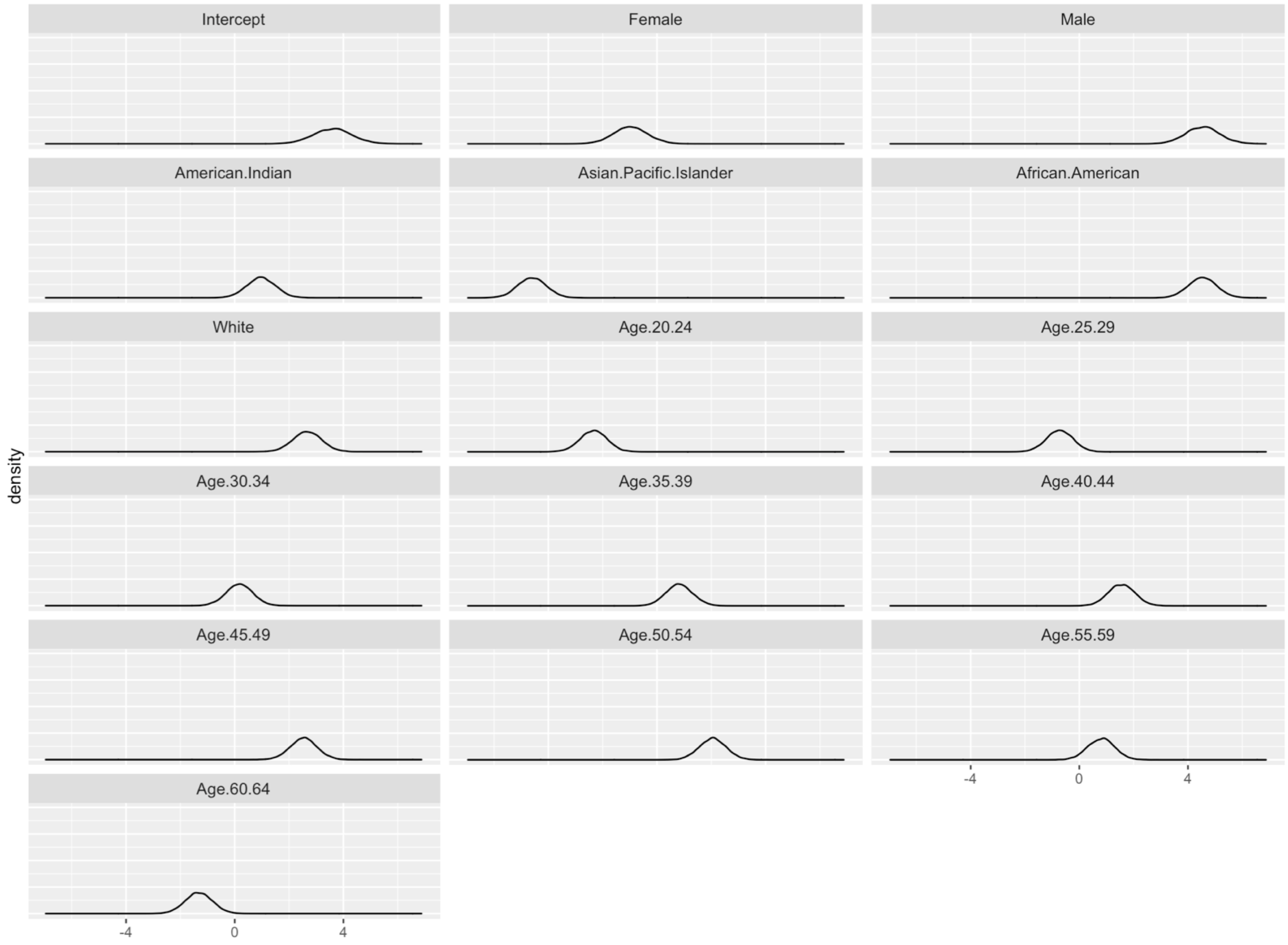

Deaths by Race
25000
20000
15000
10000
5000
0
Deaths
2000
2005
2010
2015
Year
Race
American Indian or Alaska Native
Asian or Pacific Islander
Black or African American
White


Crude Rate by Race
20
15
10
5
0
Crude Rate (per 100,000)
2000
2005
2010
2015
Year
Race
American Indian or Alaska Native
Asian or Pacific Islander
Black or African American
White

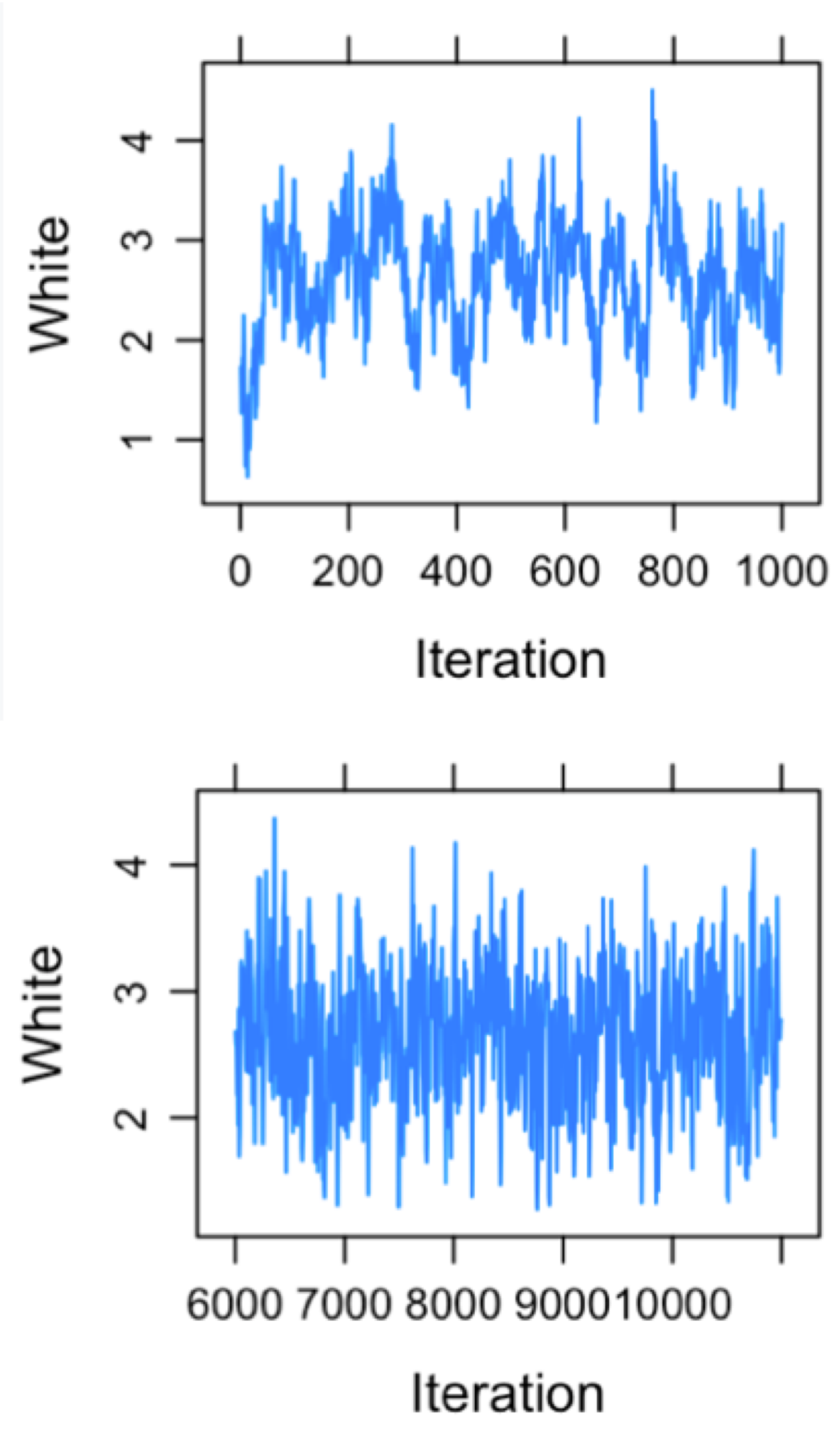
Bayesian Linear Regression

- Response variable Y_i = Crude rate
- Prior: weakly informative prior
  - $\beta \sim N(0,1)$
  - $\frac{1}{\sigma^2} \sim Gamma(1,1)$
- $Y_i^* \mid \beta_0, \beta_1, \ldots \beta_{16}, x_i, \sigma \sim N(\beta_0 + \beta_1 x_i, female + \beta_2 x_i, male + \beta_3 x_i, AsianAmerican$
$+ \beta_4 x_i, AmericanIndian + \beta_5 x_i, AfricanAmerican + \beta_6 x_i, WhiteAmerican$
$+ \beta_7 x_i, Age20-24 + \beta_8 x_i, Age25-29 + \beta_9 x_i, Age30-34 + \beta_{10} x_i, Age35-39$
$+ \beta_{11} x_i, Age40-44 + \beta_{12} x_i, Age45-49 + \beta_{13} x_i, Age50-54 + \beta_{14} x_i, Age55-59$
$+ \beta_{15} x_i, Age60-64 + \beta_{16} x_i, Time,\ \sigma)$

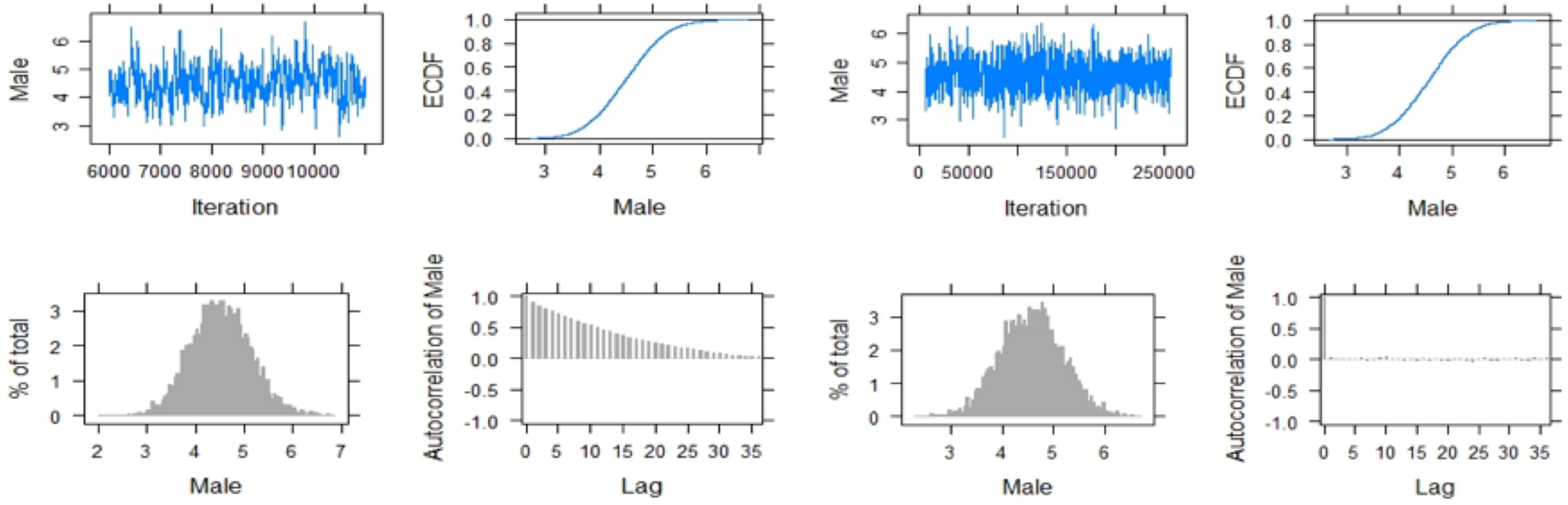
Posterior distribution for variables



Diagnostics



MCMC convergence: Traceplot of MCMC sampler for White covariate; bottom is burn-in of 5,000 samples

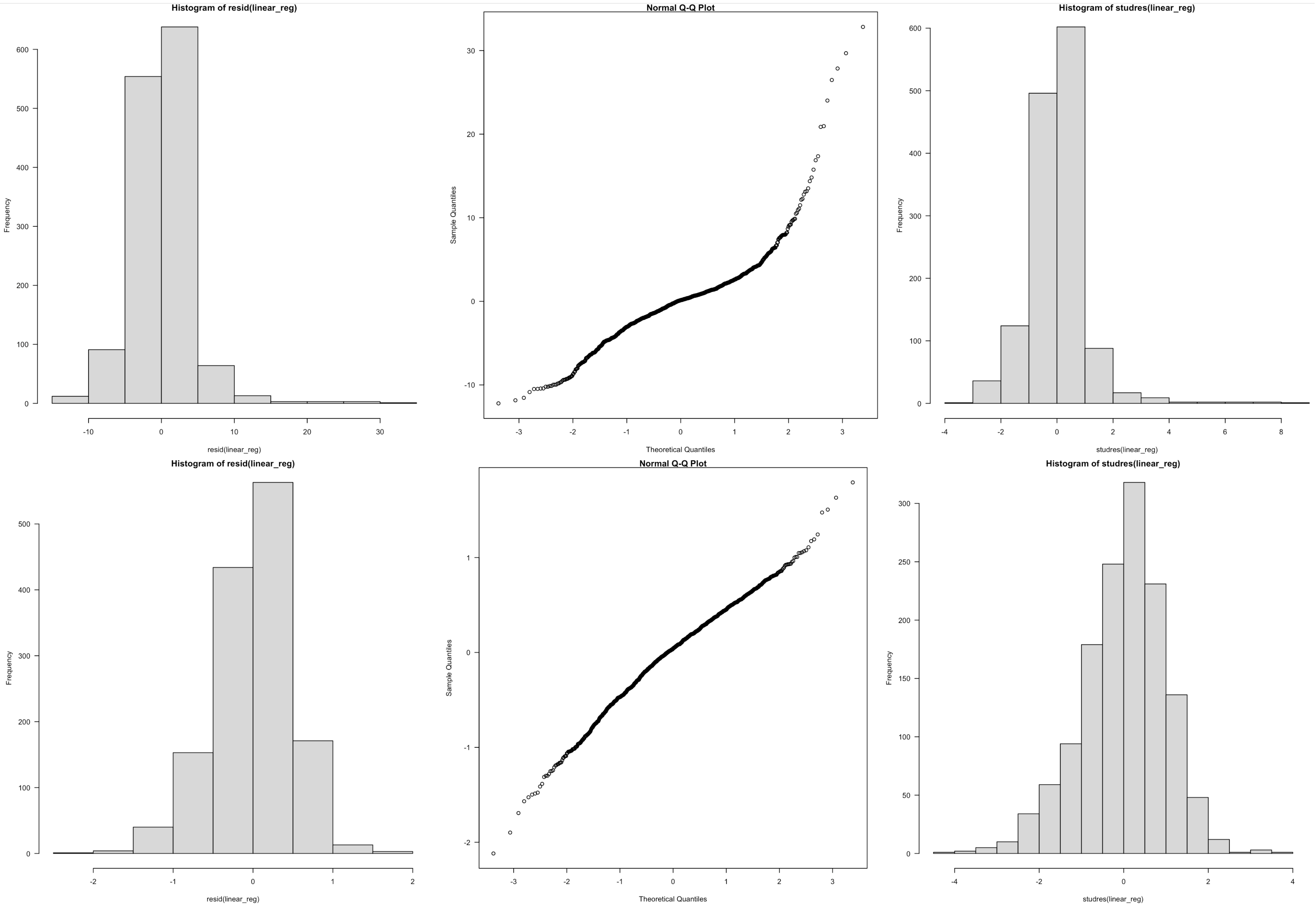


Before (left) and after (right) thinning in MCMC sampling

Multiple Linear Regression

- Response variable Y_i = Crude rate
- Factors include main effects like Year, Sex, Age, and Race
  - Interaction factors like Year with main effects and Sex with main effects
  - Quadratic time and cubic time
- In total, 55 variables
  - Most were statistically significant
  - Main effects were always larger magnitude than time effects while Sex effects had high magnitudes

Regression diagnostics



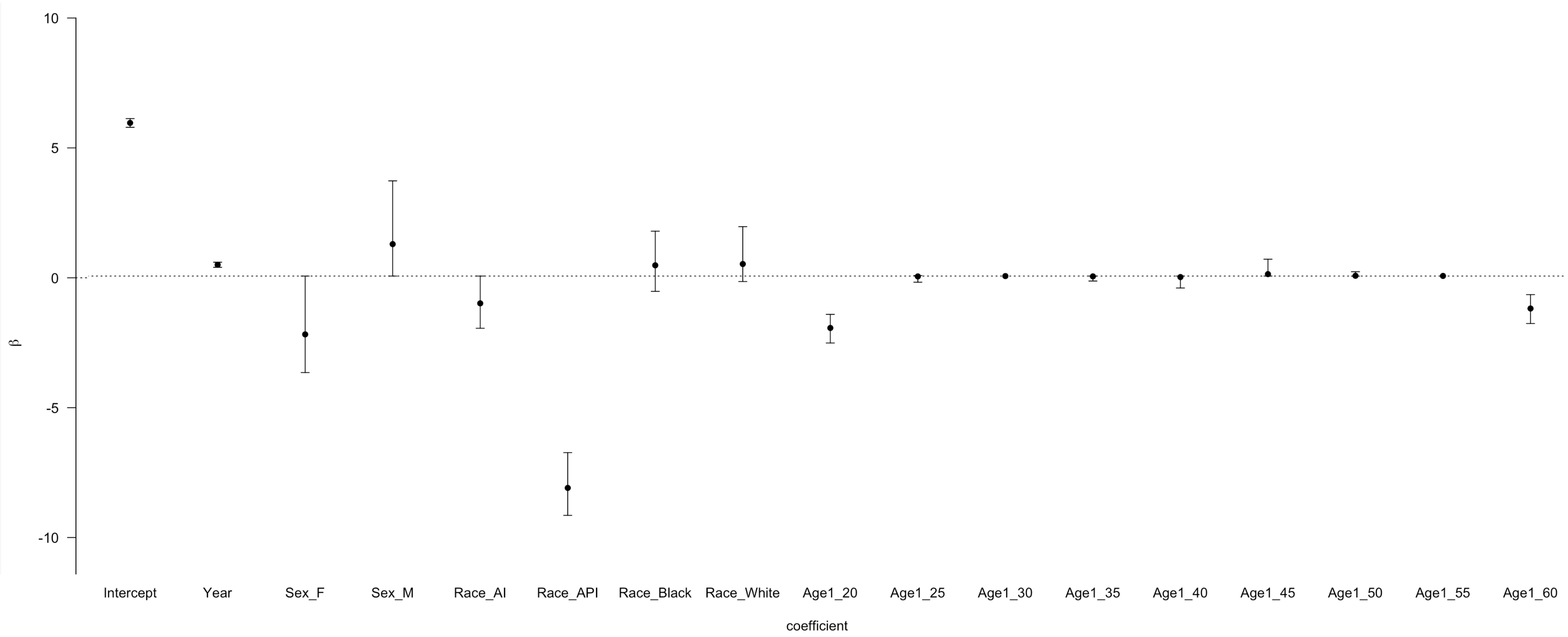
Pre-transformation
9 observations
above 4

Post-transformation
0 observations
above 4

Choosing the Best Model

- Bayesian model selection using Bayesian information criterion (BIC)
  - If model with lowest BIC is picked, may be ignoring the presence of other models that are equally good or can provide useful information
  - CI may be narrower since uncertainty is being ignored when considering only one model
- Bayesian Model Averaging: BMA
  - Uses the posterior probability of each model
  - Weighted averages of quantities of interest using these probabilities as weights (higher post. prob = higher weight)
- Stepwise selection with R squared, BIC, AIC
- LASSO penalization for variable selection

Bayesian Model Averaging

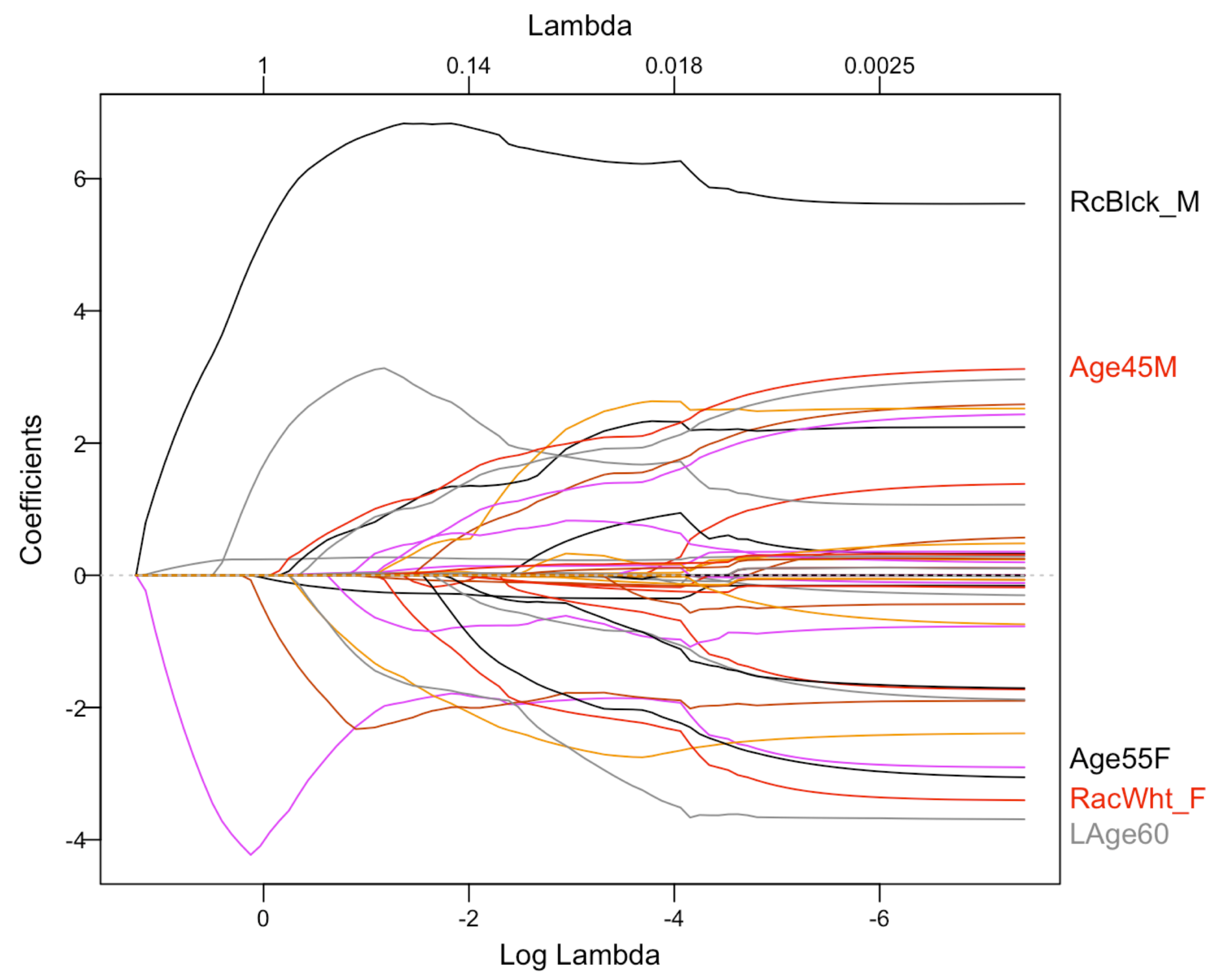


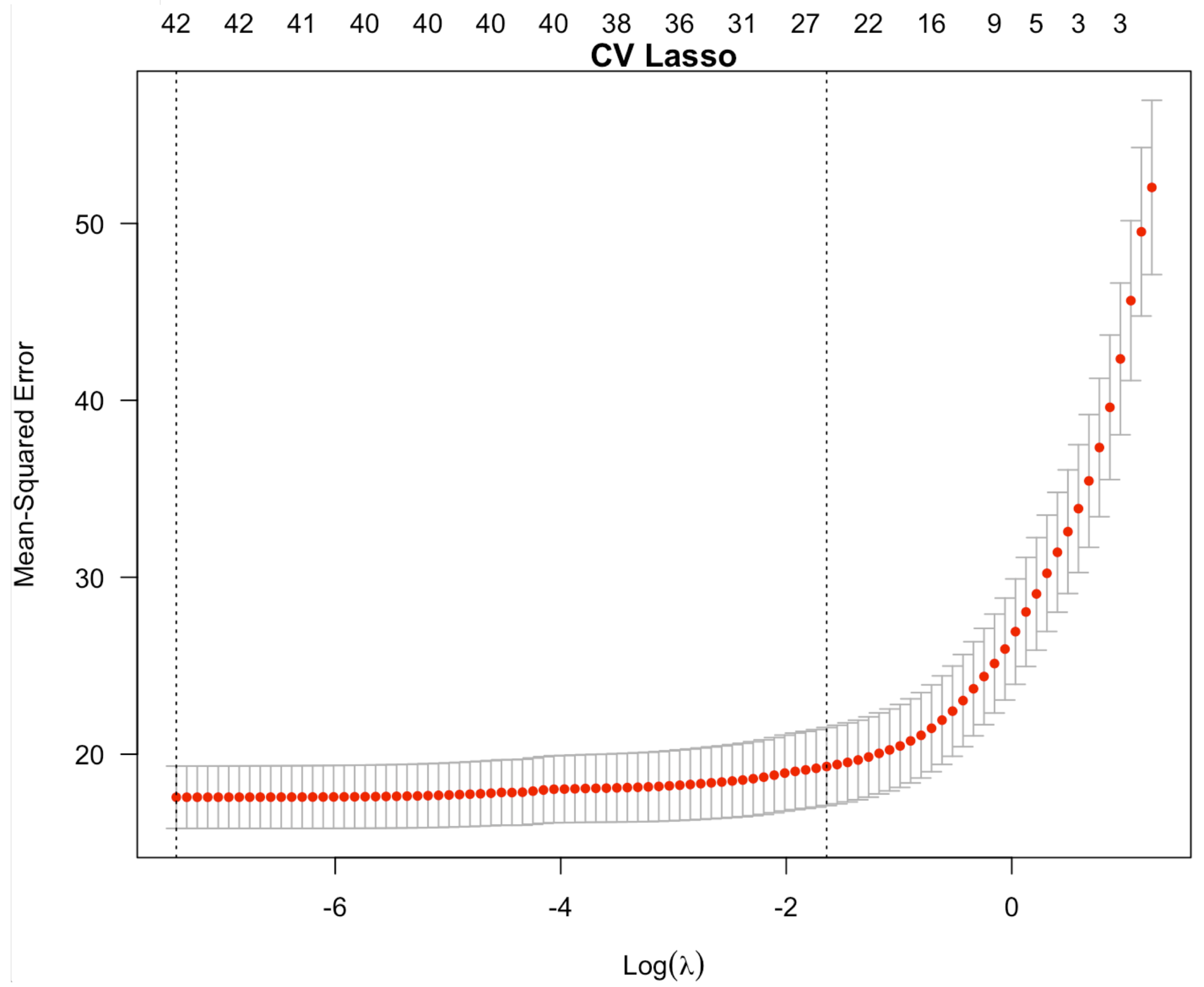
- Modeling Averaging: Year, Male, Asian, American Indian, Black, White, 20-24, 50-54, and 55-59

LASSO Penalized Regression

$$\sum_{i=1}^{n}(y_i - \sum_j x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

- Lambda 1SE vs lambda Min
  - -1.64 vs -7.41
  - 26 variables vs 43 variables
- Adjusted R-squared
  - $R^2 = 0.842$ for 43 variables
  - $R^2 = 0.823$ for 26 variables



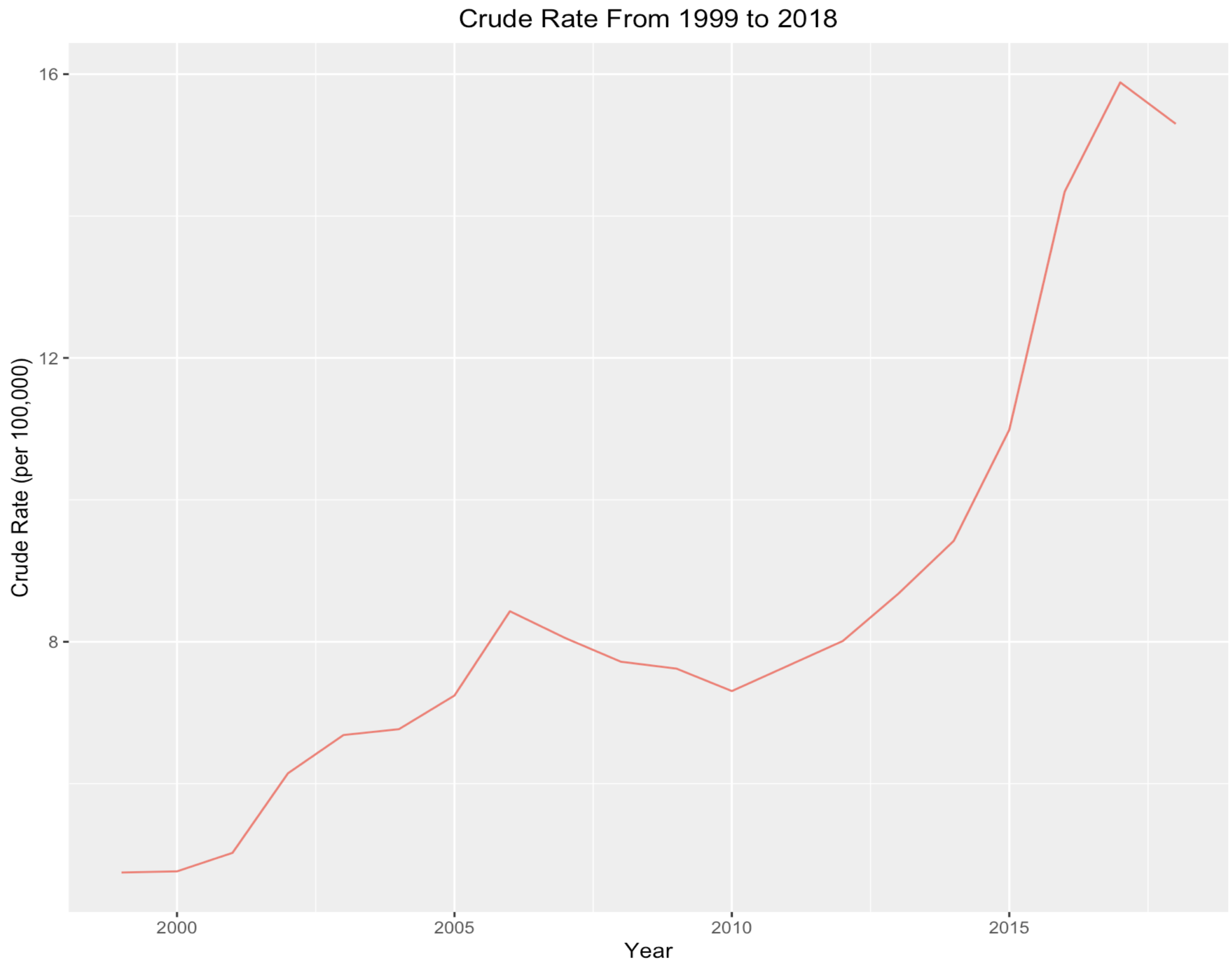


Coefficient comparison

Factor	Bayesian Model	BMA	Frequentist Model	BIC
Intercept	4.62	7.15	5.43	8.85
Time	0.13	0.22	0.09	0.11
Female	-0.97	-3.89	-5.75	-5.76
Male	4.55	1.94	3.13	3.62
Amer. Ind.	0.95	2.28	-1.84	-1.82
Asian	-4.61	-5.01	-7.89	-7.87
African	4.52	3.69	2.01	2.013
White	2.65	1.32	1.38	
20 to 24	-2.35	-0.19	-1.14	-1.56
25 to 29	-0.73		0.79	
30 to 34	0.16		1.85	1.42
35 to 39	0.822		2.61	2.18
40 to 44	1.54		3.47	3.05
45 to 49	2.51		4.62	4.2
50 to 54	2.04	-0.4	4.04	3.63
55 to 59	0.79	-1.36	2.57	2.26
60 to 64	-1.34		1.3	

Comparison of Intervals for Validation Dataset of 2017

- Bayesian model: 15 variables
- BMA Bayesian model: 13 variables
- Linear regression: 55 variables
- LASSO regression: 26 variables
- 2017 true crude rate CI: (9.10, 14.83)
  - Bayesian CI: (8.43, 11.96)
  - BMA CI: (8.13, 11.22)
  - Linear Regression CI: (8.02, 14.56)
  - LASSO CI: (7.89, 14.91)



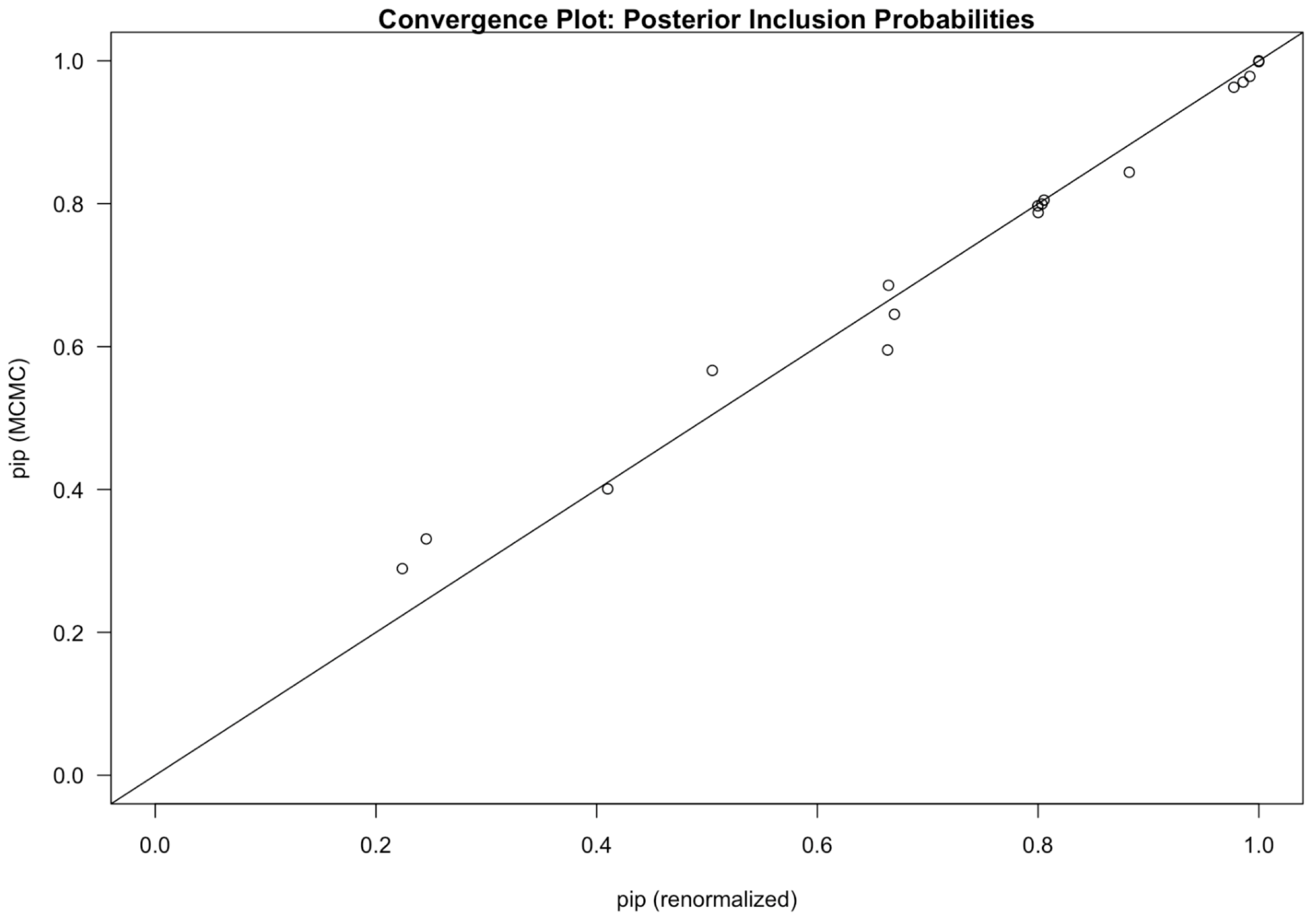
Conclusion

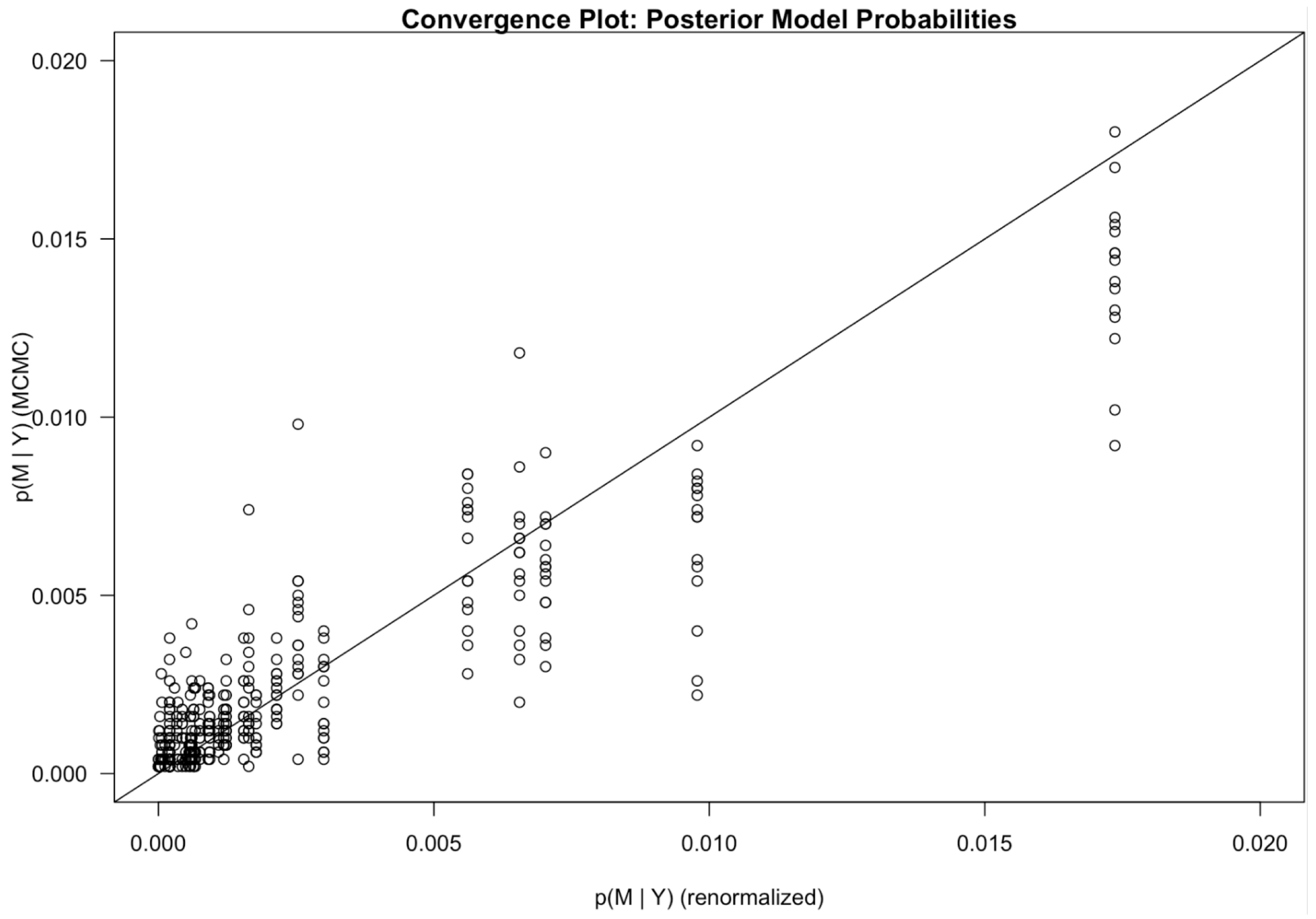
- Bayesian method provides similar variable coefficients and credible intervals while producing posterior distributions for variables

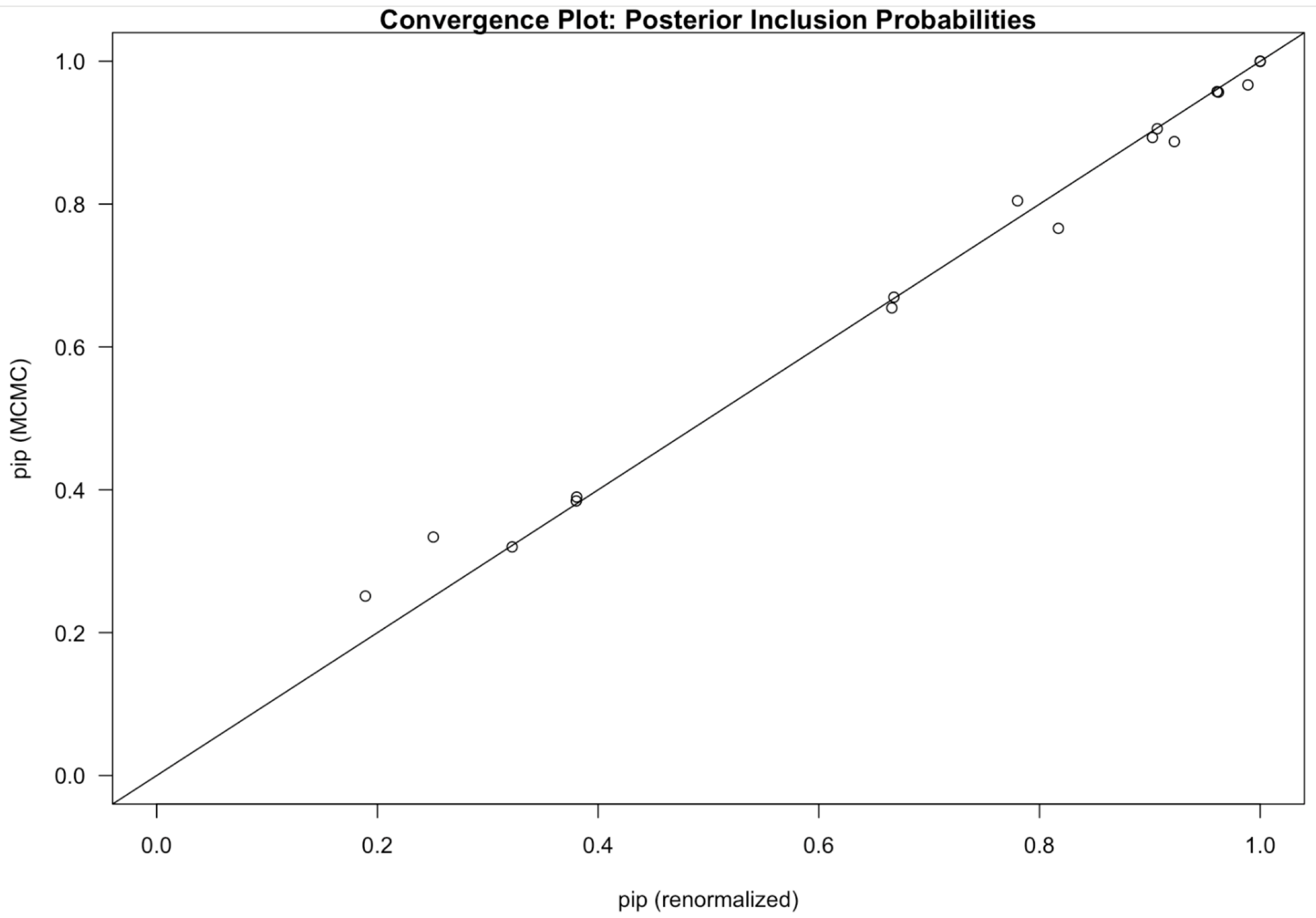
- When dealing with many interactions, LASSO is a good option for decreasing dimension space

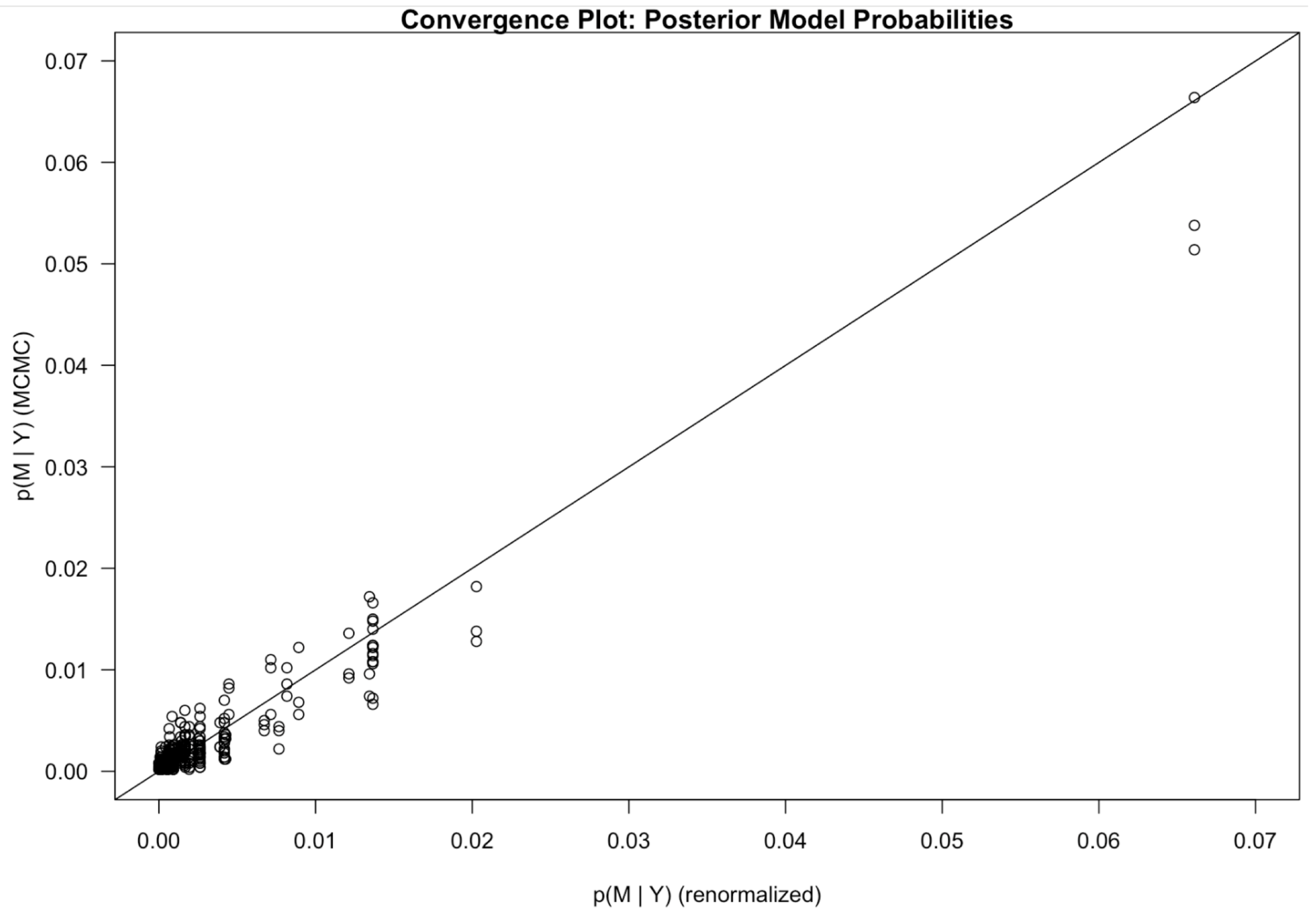
- White and African-American middle-aged males are impacted the most by the opioid crisis

Thank you


Convergence Plot: Posterior Inclusion Probabilities
pip (MCMC)
pip (renormalized)
0.0
0.2
0.4
0.6
0.8
1.0


Convergence Plot: Posterior Model Probabilities
p(M | Y) (MCMC)
p(M | Y) (renormalized)
0.000
0.005
0.010
0.015
0.020


Convergence Plot: Posterior Inclusion Probabilities
pip (MCMC)
pip (renormalized)
0.0
0.2
0.4
0.6
0.8
1.0


Convergence Plot: Posterior Model Probabilities
p(M | Y) (MCMC)
p(M | Y) (renormalized)
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07