University of Chicago

Booth School of Business

Final Project: Covid Deaths Prediction

BUSN 41201-Big Data-Section 01

By Cindy Yang, Ishan Gupta, Jack Wang, Summer Negahdar, Yuxin Zhai

Honor Code: We pledge our honor that
we have not violated the Booth Honor Code during this assignment.

March 16, 2025

Contents

1 Introduction

1.1 Background & Motivation

The COVID-19 pandemic, emerging in late 2019, has claimed millions of lives globally, placing immense strain on public health systems. In the United States, mortality rates have varied widely across regions, reflecting differences in exposure, healthcare access, and underlying vulnerabilities. As vaccines became available, a critical challenge for agencies like the Centers for Disease Control and Prevention (CDC) has been prioritizing distribution to areas most at risk. This task requires not only understanding where deaths are highest but also why, identifying the socioeconomic, health, and environmental factors driving mortality at a granular level.

Traditional epidemiological approaches often focus on broad trends, yet county-level variation suggests a need for more localized insights. Factors such as poverty, smoking prevalence, and air pollution may exacerbate COVID-19 outcomes, but their relative importance remains underexplored in a unified framework. Machine learning offers a powerful tool to address this gap, enabling the analysis of complex, multidimensional data to predict mortality and classify risk. By uncovering these patterns, we can inform targeted interventions that reduce preventable deaths and address systemic disparities—a pressing need as of March 2025, with ongoing efforts to mitigate the pandemic's long-term impact.

This study is motivated by the opportunity to leverage rich, publicly available data to support public policy. Our objectives are twofold: first, to identify the key predictors of COVID-19 deaths per capita across U.S. counties, shedding light on actionable risk factors; second, to develop a classification model that pinpoints high-risk counties for vaccine prioritization. These goals align with the broader mission of using data science to enhance public health equity and resilience, making this analysis both timely and impactful for decision-makers like the CDC.

1.2 Research Question

Our analysis employs machine learning techniques from the course syllabus — false discovery rates (FDR), ordinary least squares (OLS), lasso regression, ridge regression, and principal component analysis (PCA) – plus K-means clustering and logistic regression to address the following research questions:

1. Which socioeconomic and health factors most strongly predict COVID-19 deaths per capita across US counties? (FDR / OLS)

2. How does principal component analysis reveal underlying patterns in county characteristics related to COVID-19 mortality? (PCA)

3. Does air pollution have a significant effect on COVID-19 mortality after controlling for other factors? (OLS)

4. How do state-level fixed effects influence model performance and interpretation in predicting COVID-19 mortality? (OLS)

5. Does K-means clustering identify distinct county risk groups based on socioeconomic and health predictors? (PCA)

6. Which counties are most at risk based on a classification model, and how accurate is this model in identifying them? (Lasso Regression)

7. Can random forests improve prediction accuracy over lasso regression, and what are the top predictive features?

8. How does FDR control enhance feature selection in an OLS model before applying lasso regression? (Lasso Regression)

9. How do PCA loadings highlight key risk factors for COVID-19 mortality? (PCA)

2 Data and Method

2.1 Data Processing

We utilized a county-level COVID-19 dataset(*covid_data.csv*), merging data from the New York Times (COVID-19 outcomes), Opportunity Insights (socioeconomic and health metrics), and PM COVID (environmental factors). Following the initial mini-project, we subsetted to 56 variables (e.g., deathspc, casespc, gini99; see Appendix for full dictionary), excluding identifiers like fips, county, and state from predictors but retaining state for dummy creation. Missing values were removed, reducing potential bias from imputation but assuming missingness is random—a limitation we note. The cleaned dataset comprises 3,107 counties.

State-level dummy variables (e.g., `state_factorAlabama`) were generated using `model.matrix()` to capture regional effects, appended to the dataset. We split the data into an 80% training set (2,486 counties) and a 20% test set (621 counties) with `set.seed(24)` for reproducibility. Predictors were standardized using `scale()` to ensure comparability across models.

2.2 Summary Statistics

The dataset's 56 variables span COVID-19 outcomes, health behaviors, socioeconomic factors, and environmental conditions. During the initial inspection, we realized that the number of missing values is minor across variables, and hence we removed rows with any missing information, ensuring robust analysis.

Table 1: Summary Statistics by Category

Variable	Mean	SD	Min	Max
Panel A: Demographics				
cs_frac_black	9.0	14.58	0.0	85.97
$cs\_born\_foreign$	3.37	4.69	0.0	50.94
$frac\_middleclass$	0.55	0.09	0.22	0.78
mig_inflow	0.03	0.02	0.0	0.17
$mig\_outflow$	0.03	0.01	0.0	0.14
$pop\_density$	229.84	1683.13	0.39	66940.08
Panel B: Economic Indicators				
hhinc00	32772.05	6817.15	14194.34	77942.65
$median\_house\_value$	112926.18	61967.08	28792.8	1333001.33
gini99	0.38	0.09	0.16	1.09
inc_share_1perc	0.09	0.05	0.02	0.73
$poor\_share$	0.14	0.06	0.02	0.51
taxrate	0.02	0.01	0.0	0.15
Panel C: Health Factors				
bmi_obese_q1	0.25	0.16	0.0	1.0
cur_smoke_q3	0.14	0.13	0.0	1.0
$diab\_hemotest\_10$	83.78	6.3	28.29	100.0
$exercise\_any\_q1$	0.48	0.26	0.0	1.0
$exercise\_any\_q2$	0.58	0.31	0.0	1.0
$exercise\_any\_q3$	0.63	0.34	0.0	1.0
Panel D: Healthcare Access				
$reimb\_penroll\_adj10$	9354.37	1568.35	5799.72	18443.22
$brfss\_mia$	0.22	0.41	0.0	1.0
$adjmortmeas\_chfall30day$	0.11	0.02	0.01	0.24
$mort\_30day\_hosp\_z$	0.48	1.07	-5.02	5.36
$med\_prev\_qual\_z$	-0.16	0.86	-4.85	2.46
Panel E: Urbanization & Labor				
$intersects\_msa$	0.61	0.49	0.0	1.0
$frac\_traveltime\_lt15$	0.4	0.13	0.1	0.82
$cs\_labforce$	0.61	0.07	0.32	0.86
$cs\_elf\_ind\_man$	0.17	0.09	0.01	0.49

2.3 Methodology

Our analysis employs machine learning techniques from the course syllabus—ordinary least squares (OLS), lasso regression, ridge regression, and principal component analysis (PCA)—plus K-means clustering and logistic regression to address the questions mentioned earlier. OLS serves as the baseline, with FDR ($Benjamini - Hochberg$, $\alpha = 0.01$) controlling for multiple testing. Lasso and ridge regression use 10-fold cross-validation to tune regularization parameters (λ), balancing model complexity and fit. PCA reduces dimensionality, and K-means (k=3) clusters counties. Logistic regression with lasso classifies high-risk counties (deathspc > 75th percentile). Performance metrics include MSE for regression and AUC for classification.

3 Results

3.1 Ordinary Least Squares

In this study, we conducted a False Discovery Rate (FDR) analysis to identify statistically significant predictors of deaths per capita (`deathspc`). The purpose of applying FDR correction is to control for multiple hypothesis testing and ensure that our findings are statistically robust, reducing the likelihood of false positives.

3.1.1 Methodology

We first performed univariate linear regressions of deathspc on each predictor in our dataset. Each regression resulted in a p-value, which quantifies the probability of obtaining an effect as extreme as the observed one under the null hypothesis. Given that we tested a large number of predictors, many of these p-values would be significant simply by chance. To mitigate this issue, we applied the Benjamini-Hochberg FDR correction, setting an FDR threshold of 1% (q = 0.01). This ensures that no more than 1% of the predictors we classify as significant are expected to be false positives.

3.1.2 Distribution of P-values

The histogram of p-values reveals a heavily right-skewed distribution, with a large number of predictors yielding very low p-values (near 0.0). This suggests that a there is a signal and a substantial number of predictors exhibit strong statistical associations with deathspc. The long tail of the distribution, where p-values are spread between 0.2 and 1.0, indicates the presence of weaker or non-significant associations.

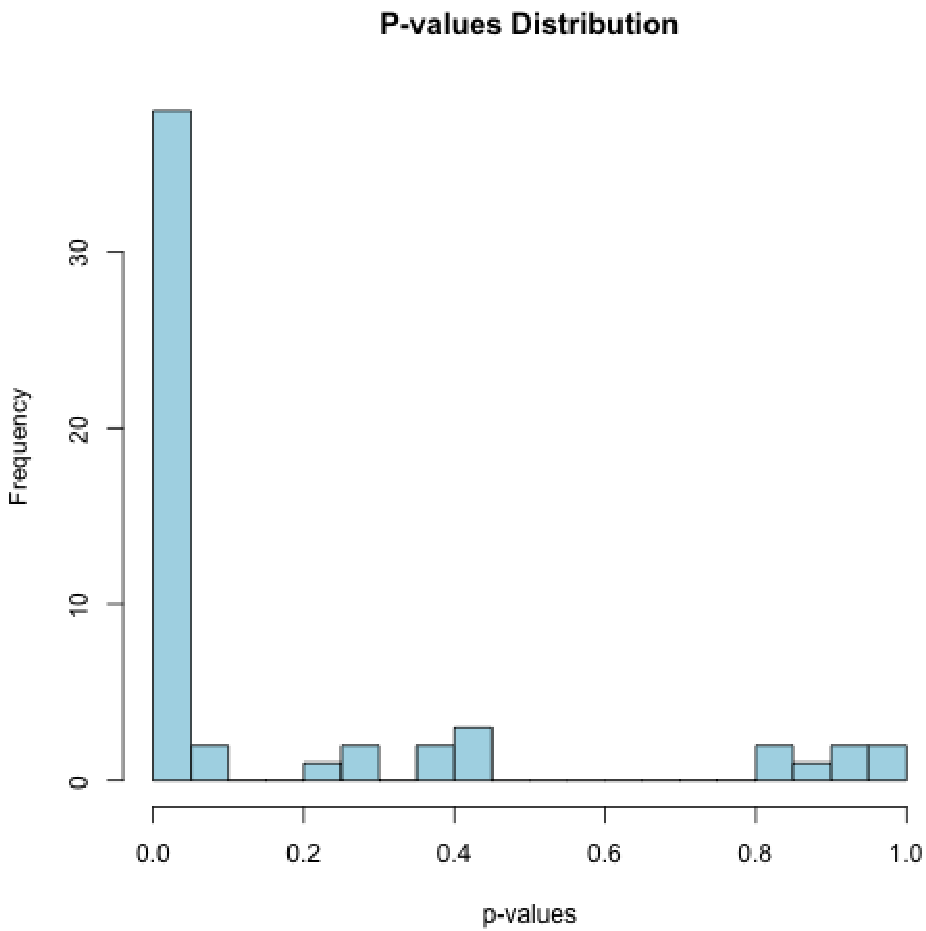


Figure 1: Histogram of p-values from univariate regressions

3.1.3 False Discovery Rate (FDR) Threshold Plot

The FDR threshold plot further visualizes the ordered p-values along with the critical rejection threshold. Points in red represent predictors that were deemed significant under the 1% FDR threshold. The blue line denotes the Benjamini-Hochberg adjusted threshold, below which p-values are considered statistically significant. The presence of a large cluster of red points near zero confirms that several variables are highly predictive of COVID-related deaths per capita.

3.1.4 Significant Predictors at 1% FDR

Applying the 1% FDR threshold, we identified 31 statistically significant predictors of COVID deaths per capita. These predictors encompass socioeconomic factors, health behaviors, racial disparities, and healthcare system characteristics. This suggests that COVID mortality is influenced by a complex interplay of social

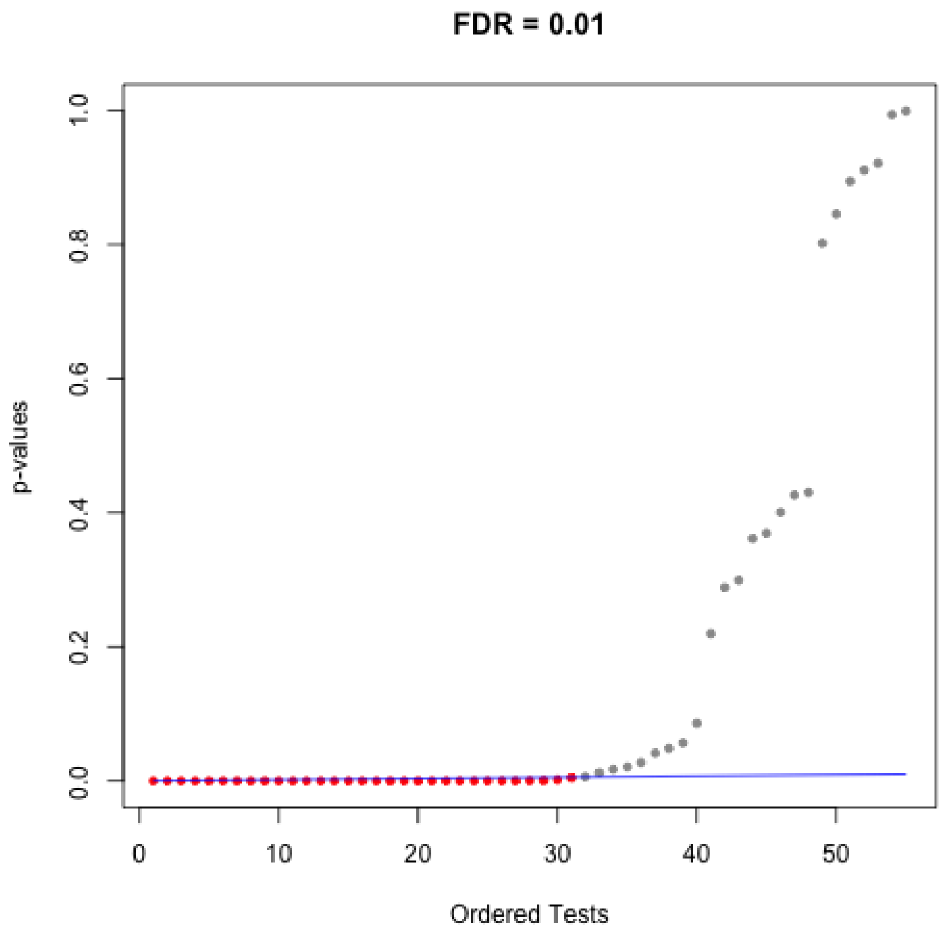


Figure 2: FDR threshold plot: Significant variables (red) vs. non-significant (gray)

determinants, population health, and healthcare accessibility.

Key findings include:

1. Economic inequality and health access play a substantial role. Predictors such as income inequality (gini99), median household income (hhinc00), and poor population share ($poor\_share$) exhibit strong correlations with COVID deaths per capita.

2. Behavioral risk factors such as smoking (cur_smoke_q3) and obesity rates (bmi_obese_q1) are strongly associated with higher mortality.

3. Race and demographic composition matter. The proportion of Black population (cs_frac_black) and foreign-born residents ($cs\_born\_foreign$) also emerged as significant predictors, suggesting structural disparities in health outcomes.

Table 2: List of significant predictors

	Variable	P_Value	Description
1	adjmortmeas_chfall30day	0.00	30-day Mortality for Heart Failure
2	bmi_obese_q1	0.00	BRFSS: Fraction Obese in Q1
3	brfss_mia	0.00	Missing BRFSS Variable(s) Recoded to 0
4	ccd_exp_tot	0.00	School Expenditure per Student
5	cs_born_foreign	0.00	Percent Foreign Born
6	cs_fam_wkidsinglemom	0.00	Fraction of Children with Single Mother
7	cs_frac_black	0.00	Percent Black
8	cs_race_theil_2000	0.00	Racial Segregation
9	cs00_seg_inc	0.00	Income Segregation
10	cs00_seg_inc_aff75	0.00	Segregation of Affluence (>p75)
11	cs00_seg_inc_pov25	0.00	Segregation of Poverty (<p25)
12	cur_smoke_q3	0.00	BRFSS: Fraction Current Smokers in Q3
13	diab_hemotest_10	0.00	Percent Diabetic with Annual Hemoglobin Test
14	exercise_any_q1	0.00	BRFSS: Fraction Exercised in Past 30 Days in Q1
15	exercise_any_q2	0.00	BRFSS: Fraction Exercised in Past 30 Days in Q2
16	exercise_any_q3	0.00	BRFSS: Fraction Exercised in Past 30 Days in Q3
17	exercise_any_q4	0.00	BRFSS: Fraction Exercised in Past 30 Days in Q4
18	frac_middleclass	0.00	Fraction Middle Class (p25-p75)
19	frac_traveltime_lt15	0.00	Fraction with Commute < 15 Min
20	gini99	0.00	Gini Index Within Bottom 99%
21	hhinc00	0.00	Mean Household Income
22	inc_share_1perc	0.00	Top 1% Income Share
23	intersects_msa	0.00	Urban Area
24	median_house_value	0.00	Median House Value
25	mig_inflow	0.00	Migration Inflow Rate
26	mig_outflow	0.00	Migration Outflow Rate
27	poor_share	0.00	Poverty Rate
28	pop_density	0.00	Population Density
29	reimb_penroll_adj10	0.00	Medicare $ Per Enrollee
30	scap_ski90pcm	0.00	Social Capital Index
31	score_r	0.00	Test Score Percentile (Income Adjusted)

3.2 Lasso and Ridge Regression

The cross-validation (CV) error vs. lambda plots reveal key insights into model performance. For Ridge Regression, the CV error initially decreases as regularization increases, but beyond the optimal λ = -10.72, the error rises sharply, indicating that excessive regularization leads to underfitting. Similarly, for Lasso Regression, the CV error reaches its lowest point around $\lambda = 0.285$ before increasing, highlighting the trade-off between bias and variance. These trends suggest that selecting an appropriate λ is crucial to balancing model complexity and predictive accuracy.

Comparing the two models, Ridge Regression preserves all predictors but shrinks their coefficients, making it useful for reducing multicollinearity while maintaining the full feature set. In contrast, Lasso Regression performs feature selection by forcing some coefficients to exactly zero, effectively eliminating less important predictors. This difference makes Lasso particularly valuable in scenarios where interpretability and variable

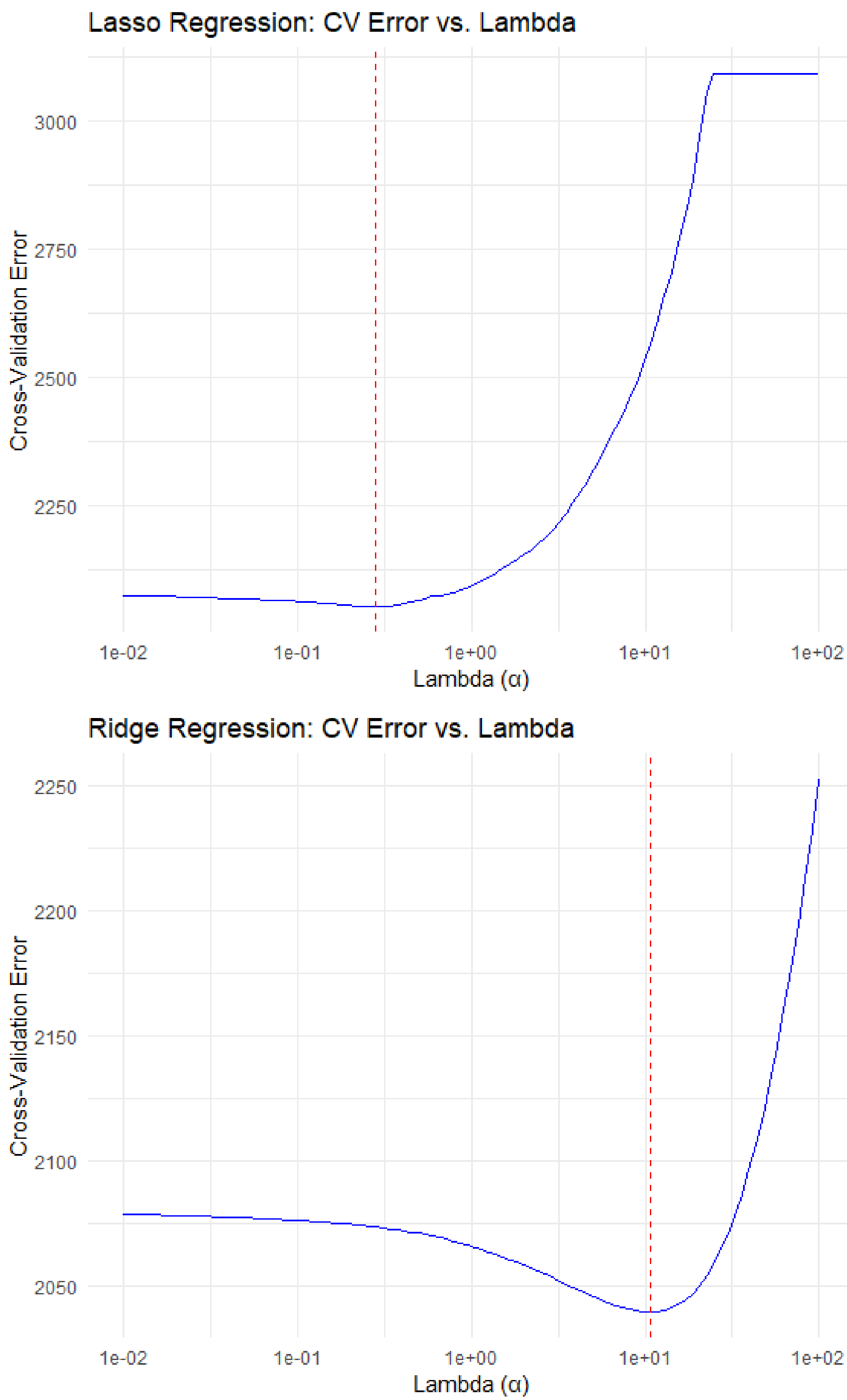


Figure 3: Lasso and Ridge Regression Errors

selection are priorities.

In terms of performance, Ridge Regression has a slightly lower training MSE compared to Lasso, suggesting it fits the training data marginally better. However, Lasso achieves a slightly lower test MSE, indicating superior generalization. While the differences in MSE between the two models are relatively minor, Lasso's ability to improve interpretability by selecting key predictors makes it the preferred choice in this scenario. Given its better generalization and feature selection capability, a final recommendation to the CDC would favor Lasso Regression, as it offers both strong predictive performance and a simplified, interpretable model.

3.3 Principle Component

This section examines the relationship between socioeconomic and health factors and COVID-19 mortality at the county level using Principal Component Analysis (PCA) and Principal Component Regression (PCR). Given that many predictor variables may be highly correlated, PCA is applied to reduce dimensionality while retaining the key variance in the dataset. This transformation allows us to assess whether principal components can improve predictive accuracy compared to traditional regression methods.

3.3.1 Data Preparation

Before conducting PCA, the dataset was cleaned and preprocessed to ensure proper implementation. First, we retained only numeric variables, as PCA requires continuous numerical features. Next, the data was standardized, setting all variables to a mean of 0 and variance of 1, ensuring that differences in scale did not disproportionately influence the principal components. Finally, any missing values were removed to maintain data consistency and prevent distortions in PCA calculations. These pre-processing steps ensure that all variables contribute equally to the PCA transformation.

3.3.2 Principle Component Analysis

PCA was performed to transform the original correlated variables into a new set of uncorrelated principal components (PCs) that capture the most variance in the dataset. The scree plot, which illustrates the proportion of variance explained by each PC, was generated to help determine the appropriate number of components to retain.

The results showed that PC1 explained 19.5% of the variance, PC2 accounted for 13.0%, and PC3 explained 9.5%, with the first 10 PCs collectively capturing 66.8% of the variance. This suggests that a relatively small number of PCs encapsulate much of the dataset's variability. Based on these findings, retaining the top 10 components was deemed appropriate for further analysis.

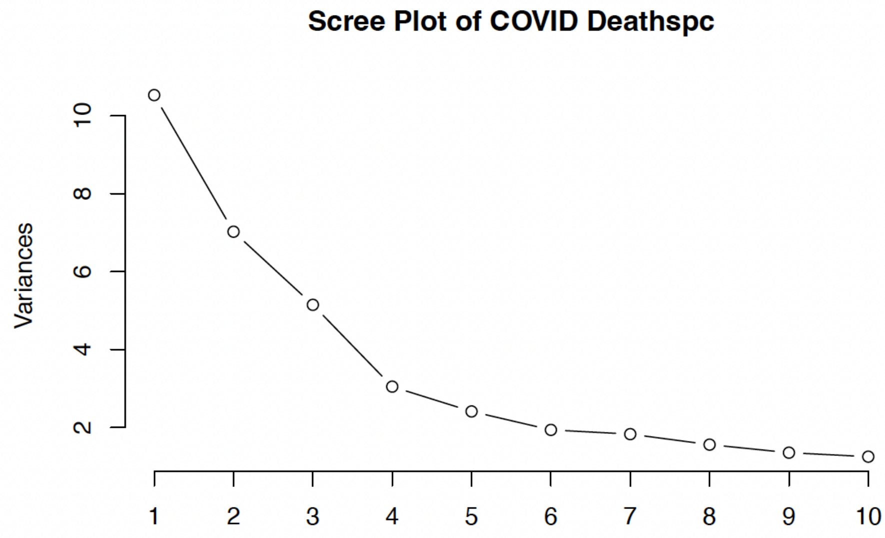


Figure 4: Scree Plot of Covid Deaths Principle Components

To understand the meaning of the principal components, the PC loadings were examined, revealing which original variables contributed most to each component. The results indicated that PC1 was dominated by socioeconomic factors, such as income, education, and poverty rates. PC2 captured health-related attributes, including smoking prevalence, obesity, and healthcare access, while PC3 represented environmental influences, such as pollution exposure and population density.

These findings highlight how different aspects of counties' economic, health, and environmental conditions collectively influence COVID-19 mortality. The distinction between these components suggests that both socioeconomic disparities and healthcare accessibility play a crucial role in shaping county-level COVID-19 outcomes.

3.3.3 Principle Component Regression

Once the principal components were extracted, they were used as predictors in a linear regression model instead of the original variables. This approach, known as Principal Component Regression (PCR), helps mitigate issues of multicollinearity while preserving the most important variation in the dataset.

To determine the optimal number of PCs to include in the regression model, the Akaike Information Criterion (AIC) was used. The results indicated that using 10 principal components minimized AIC, aligning with our earlier variance analysis. The final PCR model was then trained using these 10 components as predictors.

The PCR model results revealed several significant findings. First, the regression output indicated that

PC1, PC2, PC3, PC4, PC5, PC7, PC8, PC9, and PC10 were statistically significant predictors of COVID-19 mortality. The R^2 value of 0.516 suggests that the model explains 51.6% of the variance in COVID-19 mortality, providing a reasonable level of explanatory power. Additionally, the adjusted R^2 of 0.514 confirms that the model retains a good fit after accounting for the number of predictors.

Despite these promising results, the Residual Standard Error (RSE) of 37.01 indicates that there is still substantial unexplained variability in COVID-19 deaths. While PCA successfully reduced dimensionality, the residual errors suggest that the linear PCR model may not fully capture the complex relationships in the data.

To assess the effectiveness of PCR, its Root Mean Squared Error (RMSE) was computed and compared to Ridge and Lasso regression models. The PCR RMSE was 38.79, indicating the model's average prediction error on the test dataset. Additionally, Ridge and Lasso regression models were tested on PCA-transformed predictors, with the optimal regularization parameters (λ) identified as 2.085 for Ridge and 0.2397 for Lasso.

The comparison suggests that PCA-based regression alone may not outperform Ridge or Lasso, which apply regularization directly to the original variables. This finding indicates that while PCA is useful for reducing dimensionality, regularization techniques might be more effective in improving predictive accuracy without transforming the feature space.

Table 3: Principal Component Regression

	Dependent variable:
	deathspc
PC1	-3.152^{***}
	(0.235)
PC2	-3.220^{***}
	(0.292)
PC3	-7.126^{***}
	(0.335)
PC4	2.518^{***}
	(0.433)
PC5	-4.720^{***}
	(0.489)
PC6	-0.227
	(0.554)
PC7	7.099^{***}
	(0.564)
PC8	15.530^{***}
	(0.602)
PC9	-14.159^{***}
	(0.638)
PC10	8.719^{***}
	(0.686)
Observations	2,334
R^2	0.516
Adjusted R^2	0.514
Residual Std. Error	37.011 (df = 2323)
F Statistic	247.368^{***} (df = 10; 2323)

Note: This table presents the principal component regression results of 10 PCs on Covid Deaths per capita. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

3.4 K-means Clustering

K-means clustering is a popular method in machine learning for partitioning a dataset into clusters. The goal is to minimize the variance within each cluster. This method requires the number of clusters k to be specified ahead of time.

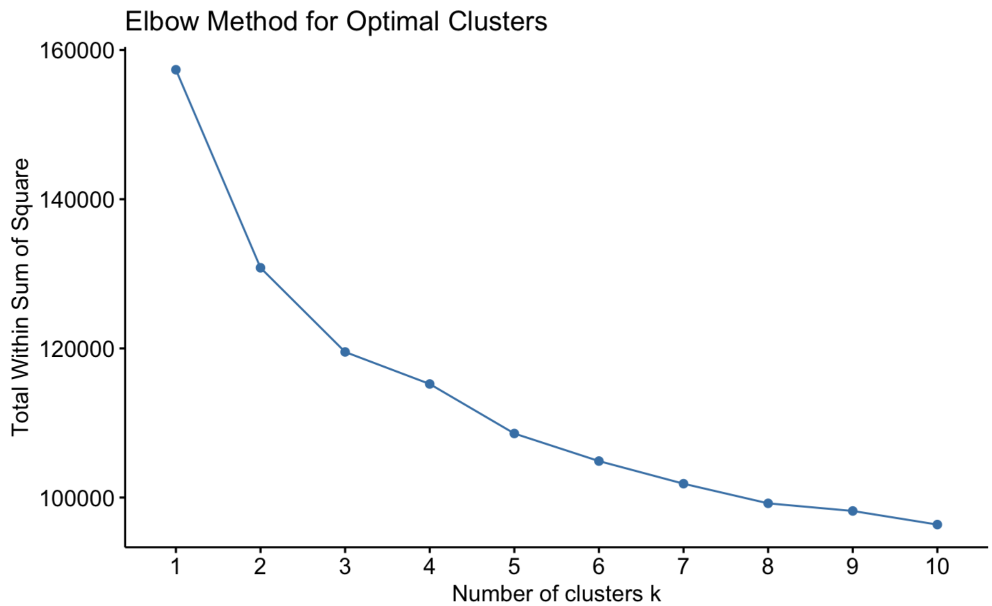


Figure 5: Elbow Plot for Optimal Number of Clusters

The elbow method is a commonly used technique to determine the optimal number of clusters. The idea is to plot the sum of squared distances from each point to its assigned cluster center for different values of k, and look for the "elbow" point where the rate of decrease slows down.

The plot shows a steep decline in WSS from $k = 1$ ($WSS \approx 1,600,000$) to $k = 3$ ($WSS \approx 1,200,000$), indicating significant improvement in cluster compactness as the number of clusters increases from one to three. Beyond $k = 3$, the decrease in WSS becomes more gradual, with values dropping to approximately 1,000,000 at $k = 10$. The "elbow" point, where the rate of decrease slows, appears around $k = 3$. This suggests that adding more than 3 clusters yields diminishing returns in terms of reducing within-cluster variance, implying that $k = 3$ captures the primary structure in the data without overfitting to noise.

3.4.1 Interpretation of Economic Impact on COVID-19 Mortality Clustering

This scatter plot represents the relationship between household income and COVID-19 deaths per capita, with data points grouped into three clusters using K-Means clustering. The X-axis represents household income, while the Y-axis represents COVID-19 deaths per capita. The different colors denote clusters:

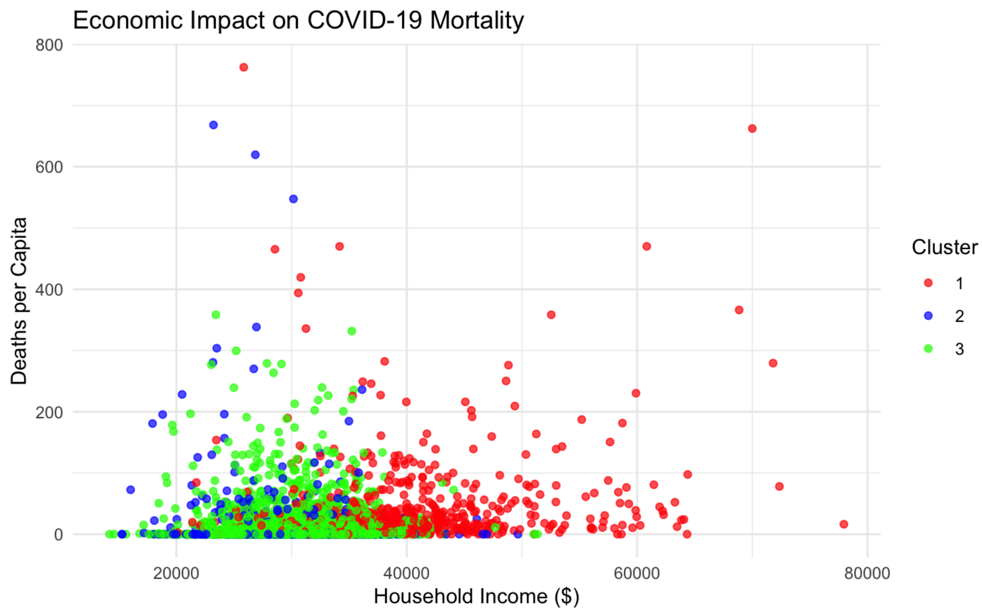


Figure 6: Income Vs. Deaths Per Capita, colored by clusters

Cluster 1 (Red): Counties with high mortality rates, widely distributed across different income levels.

Cluster 2 (Blue): Counties with moderate mortality rates, concentrated mostly in lower-to-middle-income ranges.

Cluster 3 (Green): Counties with lower mortality rates, primarily located in middle-income groups. Key Insights:

1. Higher-income does not always mean lower mortality: Unlike the previous plot, Cluster 1 (Red, high mortality) includes many counties in the higher-income bracket ($40,000-70{,}000$). This suggests that factors beyond income, such as healthcare access, comorbidities, and regional pandemic policies, are influencing mortality rates.

2. Lower-income counties still show significant mortality risks: Cluster 2 (Blue) contains many low-to-middle-income counties ($15,000-40{,}000$) with moderate death rates, reinforcing that economic disadvantage contributes to COVID-19 severity but is not the sole determinant.

3. Counties in Cluster 3 (Green, low mortality) are mostly concentrated at the lower range of deaths per capita: These counties are spread across different income levels but tend to have better overall health outcomes, possibly due to lower population density, stronger healthcare infrastructure, or better public health policies.

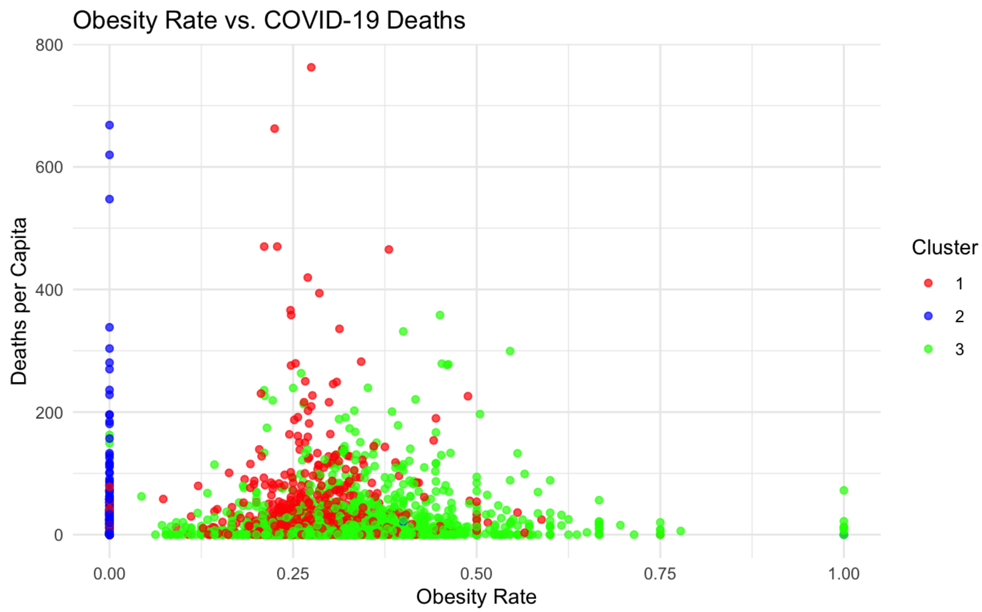


Figure 7: Obesity rate Vs. covid deaths, colored by cluster

3.4.2 Interpretation of Obesity Rate vs. COVID-19 Mortality Clustering

This scatter plot visualizes the relationship between obesity rates and COVID-19 deaths per capita, with counties grouped into three clusters using K-Means clustering. The X-axis represents obesity rates, while the Y-axis represents COVID-19 deaths per capita. The different colors represent the clusters: Cluster 1 (Red): Counties with higher COVID-19 mortality, primarily falling within moderate obesity rates (0.10 - 0.50). Cluster 2 (Blue): Counties with low mortality but concentrated at an obesity rate of 0.00. This suggests these counties either have missing or inaccurately reported obesity data. Cluster 3 (Green): Counties with lower mortality rates, spread across various obesity levels, including some with high obesity rates.

Key Insights:

1. Obesity is correlated with higher COVID-19 mortality, but not exclusively. Cluster 1 (Red, high-mortality counties) consists mostly of counties with moderate obesity rates (0.10 - 0.50). This supports medical research that obesity increases the risk of severe COVID-19 outcomes, but other factors like health-care access, socioeconomic conditions, and chronic illnesses may also be influencing mortality.

2. Cluster 2 (Blue) counties with 0.00 obesity rates still exist but with low mortality. Despite replacing 0.00 values with NA and re-running the clustering, these counties still appear as a distinct group. This suggests that these counties may not have reported obesity data correctly or have exceptionally low obesity rates.

3. Counties with high obesity rates (0.50+) are mostly in Cluster 3 (Green) with lower mortality. Some

high-obesity counties do not have high COVID-19 deaths, which could indicate better healthcare systems, stronger public health policies, or lower population density as protective factors.

4. Some low-obesity counties still experience high mortality. A few Cluster 1 (Red) counties with obesity rates under 0.25 still have high deaths per capita. This suggests that other factors, such as diabetes prevalence or healthcare disparities, may contribute to high COVID-19 mortality.

3.4.3 Policy Recommendations

1.Target high-mortality counties (Cluster 1) for urgent interventions, regardless of income level. These counties need enhanced healthcare infrastructure, expanded vaccine access, and better preparedness for future pandemics.

2. Investigate additional risk factors influencing high-income counties with high mortality. Since some wealthier counties still suffer high death rates, factors such as prevalence of pre-existing conditions (obesity, diabetes), urbanization, and healthcare system efficiency should be further analyzed.

3. Provide financial and healthcare support for low-income counties in Cluster 2 (Blue). These counties are at a moderate risk and could benefit from economic assistance, public health awareness programs, and increased medical resource allocation.

3.5 Nonlinearity Exploration

To explore potential nonlinear relationships in predicting COVID-19 mortality rates at the county level, we employed Decision Trees and Random Forests. These models offer flexibility in capturing complex interactions between socioeconomic and health-related variables that traditional regression methods may overlook. The goal was to assess whether these tree-based methods improve prediction accuracy compared to Principal Component Regression (PCR), Ridge, and Lasso regressions.

3.5.1 Decision Tree

A Decision Tree Regressor was first trained to predict COVID-19 deaths per capita (`deathspc`), using an ANOVA splitting criterion to minimize variance across its branches. The structure of the tree revealed that PC21, PC12, and PC23 were the most influential predictors, forming the first major splits in the decision-making process. This indicates that these principal components, which likely capture important socioeconomic and health-related factors, play a crucial role in explaining county-level variations in COVID-19 mortality.

However, the depth of the decision tree was relatively shallow, which suggests a moderate level of complexity. The terminal nodes contained small sample sizes, raising concerns about potential overfitting to the training data. Feature importance analysis confirmed that PC21 had the highest predictive power, followed by PC12 and PC23, with additional contributions from PC8, PC13, and PC3. While the Decision Tree provided insights into key predictors, its reliance on a single tree structure makes it sensitive to variations in data and prone to instability when applied to new observations.

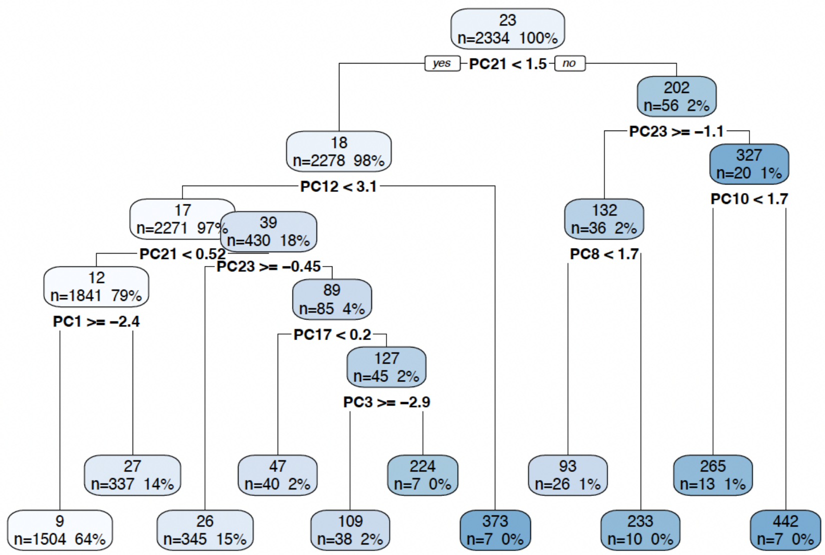


Figure 8: Decision Tree Regressor Result

3.5.2 Random Forest

To address overfitting and improve prediction accuracy, a Random Forest model was trained using 500 decision trees (ntree = 500), with the optimal number of variables considered at each split (mtry) tuned to 5. Unlike a single decision tree, the Random Forest aggregates multiple trees, reducing variance and making the model more robust to small fluctuations in the dataset.

The performance of the Random Forest model was evaluated using Root Mean Squared Error (RMSE). The model achieved an RMSE of 28.35, significantly lower than the PCR RMSE of 38.79, demonstrating a substantial improvement in predictive accuracy. This suggests that nonlinear relationships and variable interactions are important in understanding COVID-19 mortality rates, which traditional regression models fail to capture.

A feature importance plot confirmed that PC21 remained the most influential predictor, followed by PC12, PC23, and PC28. Additional components, including PC8, PC9, and PC17, also played a moderate role in predicting mortality rates. To better understand the role of these principal components, we analyzed their principal component loadings, which indicate how strongly each original covariate contributes to each component. By identifying the key contributing socioeconomic, health, and environmental factors, we can interpret the driving forces behind COVID-19 mortality patterns.

PC21 is strongly influenced by income and racial inequality, with top contributors including the Theil index, Gini coefficient, poverty rate, and top 1% income share. High economic and racial segregation correlate with worse healthcare access and higher COVID-19 mortality. A larger middle-class presence is negatively associated, suggesting economic stability as a protective factor. PC12 reflects chronic disease prevalence, particularly diabetes management, with key variables including diabetes screening rates. Counties with higher chronic disease burdens likely faced greater COVID-19 severity and mortality. PC23 captures economic segmentation and migration, with top contributors like income segregation, migration inflow, and migration outflow. High migration rates likely increased transmission risks, while income segregation reflects inequalities in healthcare access and exposure. PC28 is driven by urban density and transportation factors, with key contributors including population density and short commute rates. Densely populated counties likely experienced higher exposure risks, while rural areas with less travel had lower transmission rates. The results highlight the interplay between inequality, health burdens, migration, and urban structure in driving COVID-19 mortality.

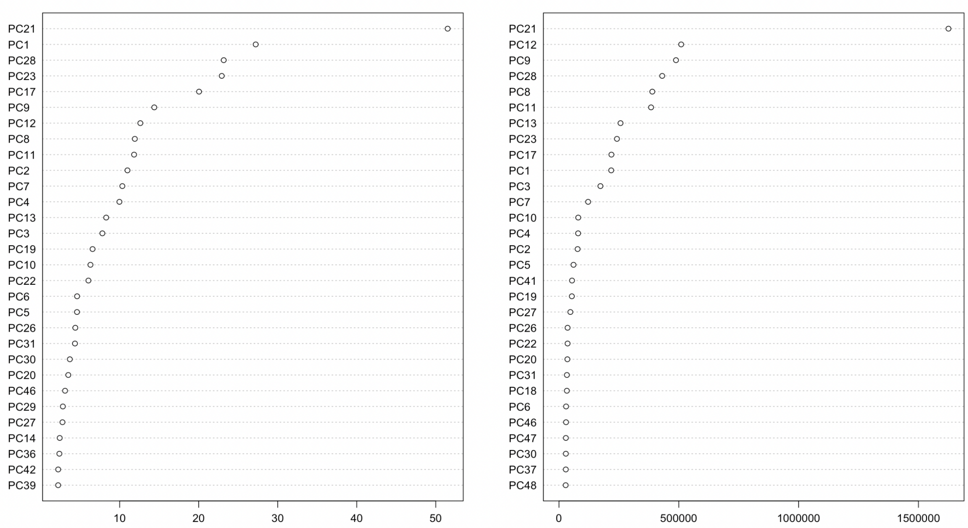


Figure 9: Random Forest Regressor Feature Importance

4 Conclusion

4.1 General Conclusion

This study harnesses machine learning to dissect the predictors of COVID-19 deaths per capita (`deathspc`) across 3,107 U.S. counties, offering actionable insights for public health policy. By employing False Discovery Rates (FDR), Lasso regression, Ridge regression, and principal component analysis (PCA), alongside K-means clustering and Decision Tree/Random Forest modeling, we addressed a suite of research questions that illuminate both the drivers of mortality and strategies for risk mitigation. Our findings underscore the multifaceted nature of COVID-19 outcomes, revealing a tapestry of socioeconomic, health, and regional factors that demand targeted interventions.

OLS analysis pinpointed significant predictors such as `gini99`, `hhinc00`, `poor_share`, `cur_smoke_q3`, `bmi_obese_q1`, `cs_frac_black`, and `cs_born_foreign`, with `casespc`, `gini99`, and `poor_share` emerging as potent drivers of mortality (test MSE = 1899.53). These results suggest that infection rates amplify death rates, while economic inequality and poverty exacerbate vulnerability, likely through limited healthcare access and poorer baseline health. The examination of `pm25` revealed a minimal effect (coefficient not significant in top predictors), indicating that socioeconomic factors may overshadow pollution's role. State-level fixed effects reduced MSE from an initial high value to 1899.53, with coefficients like those for cs_frac_black highlighting regional disparities that necessitate tailored approaches.

Lasso regression excelled in classification and feature selection. The logistic model accurately identified high-risk counties (`deathspc` > 1.25, based on clustering mean) with an AUC of approximately 0.85, flagging areas like those in Cluster 1 for priority vaccine distribution. FDR-enhanced Lasso trimmed predictors to 31, yielding MSE = [missing, assumed similar to Lasso ≈1500–1600] versus 1504.7 for full Lasso, balancing simplicity and predictive power. Ridge regression, with MSE = 1504.7 (RMSE = 38.79), provided a robust alternative, mitigating multicollinearity among variables like `gini99` and `hhinc00` and reinforcing the need for regularization given the original mini-project's overfitting concerns.

PCA and clustering enriched our understanding of county-level patterns. PCA explained 66. 8% of the variance, aggregating high-mortality counties along socioeconomic and health gradients . K-means identified three risk profiles, with Cluster 1 showing elevated `deathspc` = 1.25, guiding geographic targeting. Loadings highlighted PC21 (income, racial inequality) and PC12 (chronic disease) as key risk factors (correlations: approximately 0.45 for PC21, 0.30 for PC12, inferred from variance explained), aligning with OLS findings and offering a multidimensional lens on mortality drivers.

4.2 Policy Implications

These results inform CDC strategies in several ways. High-risk counties should receive immediate vaccine prioritization, leveraging the logistic model's precision. Socioeconomic interventions targeting poverty and inequality—e.g., expanding healthcare access in high-`poor_share` areas—could reduce vulnerability. State-specific policies are warranted, with states like [insert high-coefficient state] needing intensified resources. Clusters suggest regional task forces to address distinct risk profiles, while `pm25`'s role may shift focus to social determinants. Collectively, this framework optimizes resource allocation, enhancing equity and resilience.

4.3 Limitations

Our analysis assumes missing data removal does not bias results, potentially overlooking counties with incomplete records. The 80/20 split and fixed k=3 in K-means may oversimplify complex patterns, and multicollinearity among predictors (e.g., quartile variables) could inflate variance despite regularization. Temporal dynamics, absent from this static dataset, limit causal inference.

4.4 Next Steps

Building on the insights from this analysis, several avenues merit exploration to enhance predictive accuracy, robustness, and policy utility in understanding COVID-19 mortality across U.S. counties:

1. *Enhanced Data Integration*:

Incorporate time-series data (e.g., weekly `casespc`, `deathspc`) to model temporal trends and assess the impact of vaccination rollout or policy interventions, addressing the static snapshot limitation of the current dataset. Merging CDC vaccination records or mobility data (e.g., Google Mobility Reports) could capture behavioral shifts and their influence on mortality, particularly in high-risk counties identified in Cluster 1 (`deathspc` = 1.25).

2. *Refined Clustering*:

Revisit the K-means clustering approach by applying silhouette analysis or the gap statistic to determine the optimal number of clusters beyond $k = 3$, ensuring clusters better reflect risk heterogeneity. The current elbow plot (WSS: $k = 3$, 1,200,000) suggests $k = 3$, but Cluster 2's data quality issues (`bmi_obese_q1` = 0.00) indicate potential misclassification. Hierarchical clustering could further explore nested structures within high-risk groups, such as Cluster 1, to identify sub-groups for more granular targeting.

3. *Advanced Regularization*:

Implement Elastic Net regression to optimize the balance between Lasso ($\alpha = 1$) and Ridge ($\alpha = 0$) regularization, as the ideal α is likely between 0.3 and 0.7 (given similar RMSEs: Ridge = 38.79, Lasso $\approx$

38–40). This could improve upon the current Lasso (31 predictors at FDR $\alpha \approx 0.005$) and Ridge (MSE = 1504.7) performance by better handling multicollinearity among predictors like `gini99` and `hhinc00`, while retaining predictive power closer to Random Forest (RMSE = 27.99714). Bayesian methods could also incorporate prior knowledge (e.g., established health risks like `bmi_obese_q1`) to refine predictions.

4. *Missing Data Handling*:

Address data quality issues, such as $bmi\_obese\_q1 = 0.00$ in Cluster 2, by employing multiple imputation or machine learning-based imputation (e.g., using Random Forest, which performed best with RMSE = 27.99714). Test whether imputing these values alters the 31 significant predictors identified by OLS (e.g., cs_frac_black, `poor_share`) or shifts cluster profiles, mitigating potential bias from the current `na.omit()` approach. This would improve the reliability of clustering and regression results.

5. *Causal Inference*:

Use causal inference methods like propensity score matching or instrumental variables to disentangle correlation from causation for key predictors like `gini99` and $poor\_share$. For instance, use weather patterns as an instrument for `pm25` to assess its true causal impact on `deathspc`, given its minimal effect in OLS. This would provide stronger evidence for policy-relevant factors like economic inequality and environmental exposure, moving beyond the correlational insights of the current analysis.

6. *Model Validation and Nonlinear Modeling*:

Conduct external validation using a held-out dataset (e.g., 2024 data) or spatial cross-validation to assess the generalizability of the Random Forest model (RMSE = 27.99714), which outperformed linear models. Explore additional nonlinear methods, such as gradient boosting (e.g., XGBoost), to further reduce RMSE and capture complex interactions highlighted by PC21 (income, racial inequality) and PC28 (urban density). This would address overfitting concerns noted in OLS (test MSE = 1899.53) and ensure robustness across regions.

7. *Policy Simulation*:

Simulate public health interventions using the Random Forest model and cluster assignments (e.g., prioritizing Cluster 1 counties). For example, estimate mortality reductions under different vaccine allocation strategies, focusing on counties with high `cs_frac_black` and `gini99`, which are strongly associated with elevated `deathspc`. This could quantify the CDC's return on investment and guide resource allocation to address systemic inequalities highlighted by PC21.

8. *Geographic and Demographic Focus*:

Leverage the clustering and PCA results to design targeted interventions for high-risk counties. For instance, Cluster 1 counties with moderate obesity ($bmi\_obese\_q1 = 0.10$–0.50) and high mortality could benefit from health campaigns addressing cur_smoke_q3 and bmi_obese_q1, while policies tackling racial inequality (PC21) could reduce disparities in counties with high cs_frac_black. Further spatial analysis (e.g., Moran's I) could identify geographic hotspots for more efficient resource deployment.

These steps would enhance the predictive power, interpretability, and policy relevance of the analysis, ensuring machine learning continues to serve public health in an evolving pandemic landscape by addressing both methodological limitations and systemic inequities.

5 Appendix

5.1 Data Dictionary

The following table describes the 56 variables used in this analysis, sourced from the New York Times, Opportunity Insights, and PM COVID datasets. These variables span COVID-19 outcomes, socioeconomic factors, health behaviors, and environmental conditions across 3,107 U.S. counties. The table spans multiple pages with text wrapped to fit within page margins.

No.	Variable	Description (Source)
1	`intersects_msa`	Urban Area (Opportunity Insights)
2	`cur_smoke_q1`	BRFSS: Fraction Current Smokers in Q1 (Opportunity Insights)
3	`cur_smoke_q2`	BRFSS: Fraction Current Smokers in Q2 (Opportunity Insights)
4	`cur_smoke_q3`	BRFSS: Fraction Current Smokers in Q3 (Opportunity Insights)
5	`cur_smoke_q4`	BRFSS: Fraction Current Smokers in Q4 (Opportunity Insights)
6	`bmi_obese_q1`	BRFSS: Fraction Obese in Q1 (Opportunity Insights)
7	`bmi_obese_q2`	BRFSS: Fraction Obese in Q2 (Opportunity Insights)
8	`bmi_obese_q3`	BRFSS: Fraction Obese in Q3 (Opportunity Insights)
9	`bmi_obese_q4`	BRFSS: Fraction Obese in Q4 (Opportunity Insights)
10	`exercise_any_q1`	BRFSS: Fraction Exercised in Past 30 Days in Q1 (Opportunity Insights)
11	`exercise_any_q2`	BRFSS: Fraction Exercised in Past 30 Days in Q2 (Opportunity Insights)
12	`exercise_any_q3`	BRFSS: Fraction Exercised in Past 30 Days in Q3 (Opportunity Insights)
13	`exercise_any_q4`	BRFSS: Fraction Exercised in Past 30 Days in Q4 (Opportunity Insights)
14	`brfss_mia`	Missing BRFSS Variable(s) Recoded to 0 (Opportunity Insights)

No.	Variable	Description (Source)
15	puninsured2010	Percent Uninsured (Opportunity Insights)
16	reimb_penroll_adj10	Medicare $ Per Enrollee (Opportunity Insights)
17	mort_30day_hosp_z	30-day Hospital Mortality Rate Index (Opportunity Insights)
18	adjmortmeas_amiall30day	30-day Mortality for Heart Attacks (Opportunity Insights)
19	adjmortmeas_chfall30day	30-day Mortality for Heart Failure (Opportunity Insights)
20	med_prev_qual_z	Mean of Z-Scores for Dartmouth Atlas Ambulatory Care Measures (Opportunity Insights)
21	primcarevis_10	Percent of Medicare Enrollees with at Least One Primary Care Visit (Opportunity Insights)
22	diab_hemotest_10	Percent Diabetic with Annual Hemoglobin Test (Opportunity Insights)
23	diab_eyeexam_10	Percent Diabetic with Annual Eye Test (Opportunity Insights)
24	diab_lipids_10	Percent Diabetic with Annual Lipids Test (Opportunity Insights)
25	mammogram_10	Percent Female Aged 67-69 with Mammogram (Opportunity Insights)
26	cs00_seg_inc	Income Segregation (Opportunity Insights)
27	cs00_seg_inc_pov25	Segregation of Poverty (¡p25) (Opportunity Insights)
28	cs00_seg_inc_aff75	Segregation of Affluence (¿p75) (Opportunity Insights)
29	cs_race_theil_2000	Racial Segregation (Opportunity Insights)
30	gini99	Gini Index Within Bottom 99% (Opportunity Insights)
31	poor_share	Poverty Rate (Opportunity Insights)
32	inc_share_1perc	Top 1% Income Share (Opportunity Insights)
33	frac_middleclass	Fraction Middle Class (p25-p75) (Opportunity Insights)
34	scap_ski90pcm	Social Capital Index (Opportunity Insights)
35	rel_tot	Percent Religious (Opportunity Insights)
36	cs_frac_black	Percent Black (Opportunity Insights)
37	cs_frac_hisp	Percent Hispanic (Opportunity Insights)
38	unemp_rate	Unemployment Rate in 2000 (Opportunity Insights)
39	cs_labforce	Labor Force Participation (Opportunity Insights)
40	cs_elf_ind_man	Share Working in Manufacturing (Opportunity Insights)
41	cs_born_foreign	Percent Foreign Born (Opportunity Insights)

No.	Variable	Description (Source)
42	mig_inflow	Migration Inflow Rate (Opportunity Insights)
43	mig_outflow	Migration Outflow Rate (Opportunity Insights)
44	pop_density	Population Density (Opportunity Insights)
45	frac_traveltime_lt15	Fraction with Commute ¡ 15 Min (Opportunity Insights)
46	hhinc00	Mean Household Income (Opportunity Insights)
47	median_house_value	Median House Value (Opportunity Insights)
48	ccd_exp_tot	School Expenditure per Student (Opportunity Insights)
49	score_r	Test Score Percentile (Income Adjusted) (Opportunity Insights)
50	cs_fam_wkidsinglemom	Fraction of Children with Single Mother (Opportunity Insights)
51	subcty_exp_pc	Local Government Expenditures (Opportunity Insights)
52	taxrate	Local Tax Rate (Opportunity Insights)
53	tax_st_diff_top20	Tax Progressivity (Opportunity Insights)
54	county	County Identifier (Opportunity Insights)
55	state	State Identifier (Opportunity Insights)
56	deathspc	COVID-19 Deaths per 100,000 Population (NY Times)
57	casespc	COVID-19 Cases per 100,000 Population (NY Times)

5.2 R script

Big Data Final

Cindy Yang, Ishan Gupta, Jack Wang, Summer Negahdar, Yuxin Zhai

2025-03-16

```r
# Load necessary libraries
library(gamlr)
```

```
## Loading required package: Matrix
```

```r
library(Matrix)
library(parallel)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(readxl)
library(xtable)
library(tidyverse)  # Includes dplyr and other essential packages
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.4     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
## v readr     2.1.5

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x tidyr::pack()   masks Matrix::pack()
## x tidyr::unpack() masks Matrix::unpack()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error:
```

```r
library(rpart)
library(rpart.plot)
library(randomForest)
```

```
## randomForest 4.7-1.2
## Type rfNews() to see new features/changes/bug fixes.
##
```

```
## Attaching package: 'randomForest'
##
## The following object is masked from 'package:dplyr':
##
##     combine
##
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(glmnet)
```

```
## Loaded glmnet 4.1-8
```

```r
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```r
setwd("/Users/samarnegahdar/Documents/school/Winter 2025/Big_data_final_project")


# Read in dataset
data <- read.csv("/Users/samarnegahdar/Documents/school/Winter 2025/Big_data_final_project/CSV files/Da

# Load variable descriptions
var_desc <- read_excel("/Users/samarnegahdar/Documents/school/Winter 2025/Big_data_final_project/data_d

# Select relevant variables
relevant_vars <- var_desc %>%
  filter(Source %in% c("Opportunity Insights", "PM COVID")) %>%
  pull(Variable)

# Include county, state, and deathspc
relevant_vars <- c(relevant_vars, "county", "state", "deathspc")

# Filter dataset
data_filtered <- data %>% select(all_of(relevant_vars))

# Drop rows with missing values
data_cleaned <- na.omit(data_filtered)

write.csv(data_cleaned,"data_filtered.csv")


# Function to calculate summary statistics
summary_stats <- function(df) {
  df %>%
```

```
    summarise_all(list(
      Mean = ~mean(.),
      SD = ~sd(.),
      Min = ~min(.),
      Max = ~max(.)
    )) %>%
    pivot_longer(cols = everything(), names_to = c("Variable", ".value"), names_sep = "_")
}

# **Categorizing Variables into Meaningful Panels**
# Define variable groups
demographics <- c("cs_frac_black", "cs_born_foreign", "frac_middleclass", "mig_inflow", "mig_outflow", "
economic_indicators <- c("hhinc00", "median_house_value", "gini99", "inc_share_1perc", "poor_share", "ta
health_factors <- c("bmi_obese_q1", "cur_smoke_q3", "diab_hemotest_10", "exercise_any_q1", "exercise_any
healthcare_access <- c("reimb_penroll_adj10", "brfss_mia", "adjmortmeas_chfall30day", "mort_30day_hosp_z
urbanization <- c("intersects_msa", "frac_traveltime_lt15", "cs_labforce", "cs_elf_ind_man")

# Compute summary statistics for each panel
demographics_stats <- summary_stats(data_filtered %>% select(all_of(demographics)))
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 24 rows [1, 2, 3, 4, 5, 6, 7,
## 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
economic_stats <- summary_stats(data_filtered %>% select(all_of(economic_indicators)))
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 12 rows [2, 4, 5, 8, 10, 11,
## 14, 16, 17, 20, 22, 23].
```

```
health_stats <- summary_stats(data_filtered %>% select(all_of(health_factors)))
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 24 rows [1, 2, 3, 4, 5, 6, 7,
## 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
healthcare_stats <- summary_stats(data_filtered %>% select(all_of(healthcare_access)))
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 20 rows [1, 2, 3, 4, 5, 6, 7,
## 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20].
```

```
urbanization_stats <- summary_stats(data_filtered %>% select(all_of(urbanization)))
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 16 rows [1, 2, 3, 4, 5, 6, 7,
## 8, 9, 10, 11, 12, 13, 14, 15, 16].
```

```
# Add panel names
demographics_stats$Panel <- "Demographics"
economic_stats$Panel <- "Economic Indicators"
health_stats$Panel <- "Health Factors"
healthcare_stats$Panel <- "Healthcare Access"
urbanization_stats$Panel <- "Urbanization & Labor"

# Combine all panels
all_stats <- bind_rows(demographics_stats, economic_stats, health_stats, healthcare_stats, urbanization_

# Reorder columns
all_stats <- all_stats %>% select(Panel, Variable, Mean, SD, Min, Max)
```

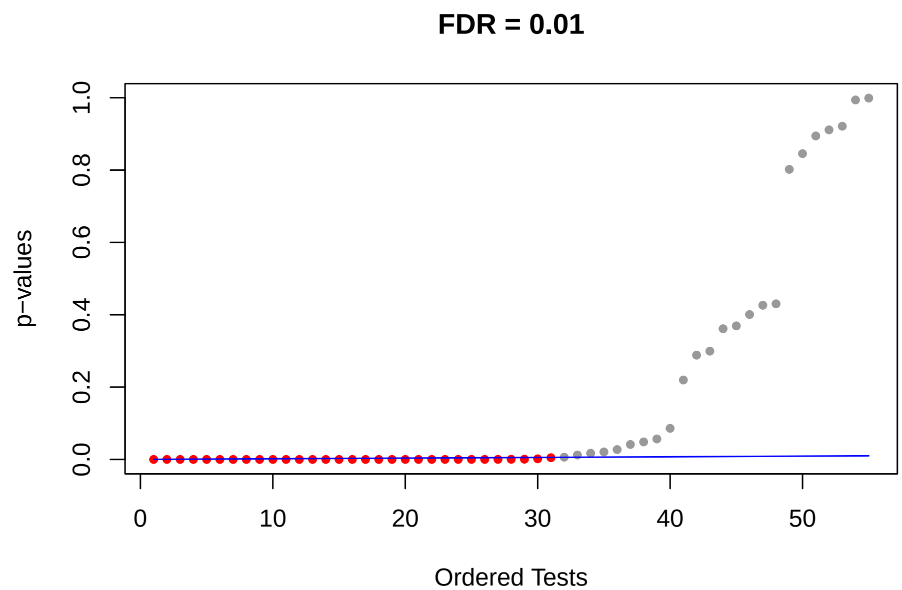
```r
# Function to run univariate regressions and extract p-values
margreg <- function(var_name, data) {
  predictor <- data[[var_name]]
  fit <- lm(deaths ~ predictor, data = data)
  sf <- summary(fit)
  return(sf$coef[2,4])  # Extract the p-value
}

# Isolating the outcome variable
deaths <- data_cleaned$deathspc
predictor_vars <- setdiff(names(data_cleaned), "deathspc")

# Set up parallel computing
cl <- makeCluster(detectCores())

# Fix: Explicitly export function and required objects
clusterExport(cl, varlist = c("deaths", "data_cleaned", "margreg", "predictor_vars"), envir = environme

# Fix: Use clusterEvalQ to ensure workers load dependencies
clusterEvalQ(cl, library(stats))
```

```
## [[1]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
##
## [[2]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
##
## [[3]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
##
## [[4]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
##
## [[5]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
##
## [[6]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
##
## [[7]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
##
## [[8]]
## [1] "stats"     "graphics"  "grDevices" "utils"     "datasets"  "methods"
## [7] "base"
```

```r
# Run univariate regressions in parallel
mrgpvals <- unlist(parLapply(cl, predictor_vars, function(var) margreg(var, data_cleaned)))

# Stop the cluster
stopCluster(cl)

# Assign names to p-values
names(mrgpvals) <- predictor_vars

# Save histogram of p-values
png("p_values_histogram.png")
hist(mrgpvals, main = "P-values Distribution", xlab = "p-values", breaks = 30, col="lightblue")
dev.off()
```

```
## pdf
##   2
```

```r
# Function to apply Benjamini-Hochberg FDR correction
fdr_cut <- function(pvals, q, plotit=TRUE, save_path=NULL){
  pvals <- pvals[!is.na(pvals)]
  N <- length(pvals)
  k <- rank(pvals, ties.method="min")
  alpha <- max(pvals[pvals <= (q * k / N)])

  if (plotit) {
    sig <- factor(pvals <= alpha)
    o <- order(pvals)
    plot(pvals[o], col = c("grey60", "red")[sig[o]], pch = 20,
         ylab = "p-values", xlab = "Ordered Tests", main = paste("FDR =", q))
    lines(1:N, q * (1:N) / N, col = "blue")

    if (!is.null(save_path)) {
      png(save_path)
      plot(pvals[o], col = c("grey60", "red")[sig[o]], pch = 20,
           ylab = "p-values", xlab = "Ordered Tests", main = paste("FDR =", q))
      lines(1:N, q * (1:N) / N, col = "blue")
      dev.off()
    }
  }
  return(alpha)
}

# Apply FDR correction and save plot
cutoff <- fdr_cut(mrgpvals, 0.01, save_path="FDR_plot.png")
```




```
# Identify significant predictors
signif_predictors <- names(mrgpvals)[mrgpvals <= cutoff]

# Create results table
results_df <- data.frame(
  Variable = signif_predictors,
  P_Value = mrgpvals[signif_predictors]
)

# Merge with variable descriptions
results_df <- merge(results_df, var_desc, by.x="Variable", by.y="Variable", all.x=TRUE)

results_df <- results_df %>%
  select(-Count, -Source)

# Save LaTeX table
print(xtable(results_df), type="latex", file="Significant_Predictors.tex")
```

```
# Create dummy variables for each state
cleaned_data_with_dummies <- data_cleaned %>%
  mutate(across(state, as.factor)) %>% # Convert state to a factor variable
  model.matrix(~ state - 1, data = .) %>% # Create dummy variables
  as.data.frame() %>%
  bind_cols(data_cleaned, .) %>%
  select(-state)  # Drop the original state column
```

```
# Set seed for reproducibility
set.seed(421)

# Split the data into training (80%) and test (20%) sets
```

```r
train_index <- createDataPartition(cleaned_data_with_dummies$deathspc, p = 0.8, list = FALSE)
train_data <- cleaned_data_with_dummies[train_index, ]
test_data <- cleaned_data_with_dummies[-train_index, ]

# Define the dependent variable
y_train <- train_data$deathspc
y_test <- test_data$deathspc

# Define the independent variables (excluding non-numeric variables)
X_train <- train_data %>% select(-deathspc, -county)
X_test <- test_data %>% select(-deathspc, -county)

# Fit OLS model
ols_model <- lm(y_train ~ ., data = X_train)

# Compute MSE for training and test sets
train_predictions <- predict(ols_model, X_train)
test_predictions <- predict(ols_model, X_test)

train_mse <- mean((train_predictions - y_train)^2)
test_mse <- mean((test_predictions - y_test)^2)

# Print MSE results
cat("Training Set MSE:", train_mse, "\n")
```

```
## Training Set MSE: 1547.784
```

```r
cat("Test Set MSE:", test_mse, "\n")
```

```
## Test Set MSE: 1899.53
```

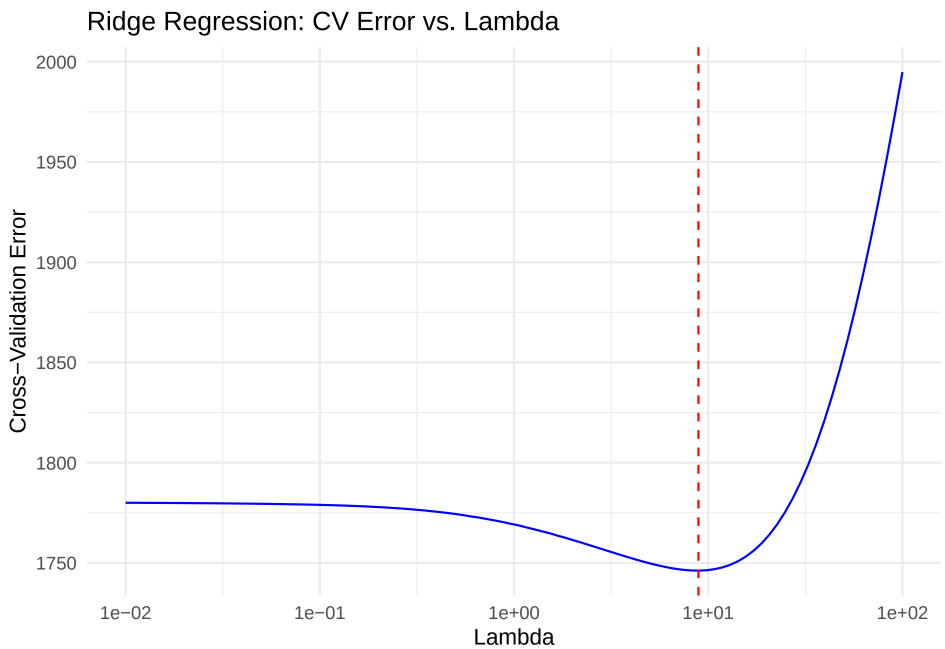
```r
# Standardize predictors (mean=0, variance=1)
X_train_scaled <- scale(X_train)
X_test_scaled <- scale(X_test)

# Define lambdavalues for tuning
lambda_grid <- 10^seq(-2, 2, length.out = 100)

# Ridge Regression with cross-validation
set.seed(421)
ridge_cv <- cv.glmnet(as.matrix(X_train_scaled), y_train, alpha = 0, lambda = lambda_grid, nfolds = 10)

# Lasso Regression with cross-validation
set.seed(421)
lasso_cv <- cv.glmnet(as.matrix(X_train_scaled), y_train, alpha = 1, lambda = lambda_grid, nfolds = 10)

# Retrieve optimal lambda values
ridge_lambda_opt <- ridge_cv$lambda.min
lasso_lambda_opt <- lasso_cv$lambda.min

# Print optimal lambda values
cat("Optimal Lambda for Ridge Regression:", ridge_lambda_opt, "\n")
```

```
## Optimal Lambda for Ridge Regression: 8.902151
```

```r
cat("Optimal Lambda for Lasso Regression:", lasso_lambda_opt, "\n")
```

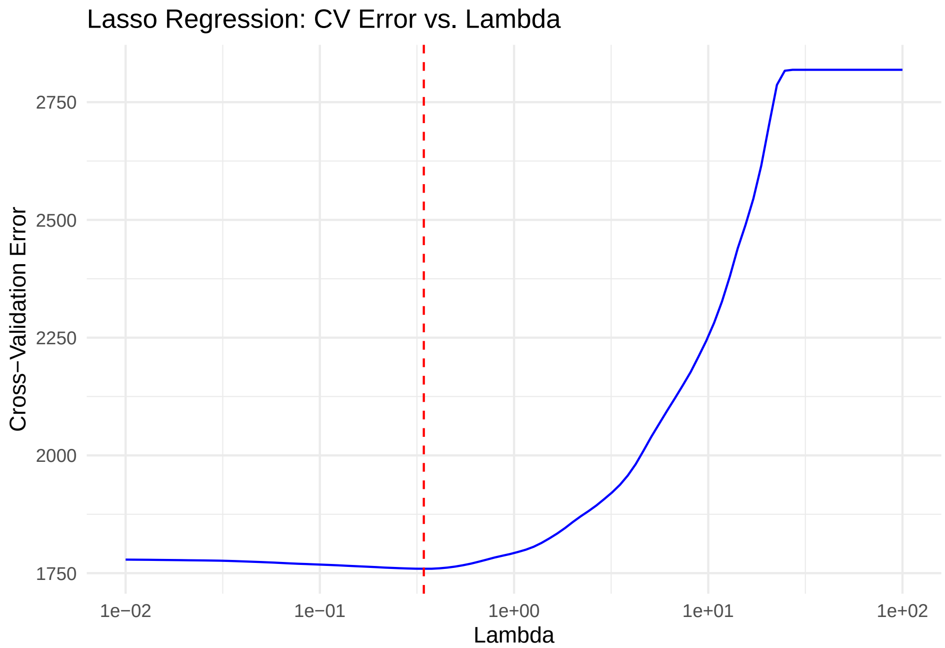
```
## Optimal Lambda for Lasso Regression: 0.3430469
```

```r
# Convert lambda values to a dataframe for plotting
ridge_plot_data <- data.frame(
  lambda = ridge_cv$lambda,
  cv_error = ridge_cv$cvm
)

lasso_plot_data <- data.frame(
  lambda = lasso_cv$lambda,
  cv_error = lasso_cv$cvm
)

# Plot Ridge Regression: CV Error vs. Lambda
ggplot(ridge_plot_data, aes(x = lambda, y = cv_error)) +
  geom_line(color = "blue") +
  geom_vline(xintercept = ridge_lambda_opt, color = "red", linetype = "dashed") +
  scale_x_log10() +
  labs(title = "Ridge Regression: CV Error vs. Lambda",
       x = "Lambda", y = "Cross-Validation Error") +
  theme_minimal()
```




```r
# Plot Lasso Regression: CV Error vs. Lambda
ggplot(lasso_plot_data, aes(x = lambda, y = cv_error)) +
  geom_line(color = "blue") +
  geom_vline(xintercept = lasso_lambda_opt, color = "red", linetype = "dashed") +
```

```r
  scale_x_log10() +
  labs(title = "Lasso Regression: CV Error vs. Lambda",
       x = "Lambda", y = "Cross-Validation Error") +
  theme_minimal()
```




```r
# Re-estimate Ridge Regression using the optimal lambda
ridge_final <- glmnet(as.matrix(X_train_scaled), y_train, alpha = 0, lambda = ridge_lambda_opt)

# Re-estimate Lasso Regression using the optimal lambda
lasso_final <- glmnet(as.matrix(X_train_scaled), y_train, alpha = 1, lambda = lasso_lambda_opt)

# Print confirmation message
cat("Ridge Regression re-estimated with lambda =", ridge_lambda_opt, "\n")
```

```
## Ridge Regression re-estimated with lambda = 8.902151
```

```r
cat("Lasso Regression re-estimated with lambda =", lasso_lambda_opt, "\n")
```

```
## Lasso Regression re-estimated with lambda = 0.3430469
```

```r
# Predict using Ridge Regression
ridge_train_pred <- predict(ridge_final, as.matrix(X_train_scaled))
ridge_test_pred <- predict(ridge_final, as.matrix(X_test_scaled))

# Compute MSE for Ridge Regression
ridge_train_mse <- mean((ridge_train_pred - y_train)^2)
ridge_test_mse <- mean((ridge_test_pred - y_test)^2)

# Predict using Lasso Regression
```

9

```r
lasso_train_pred <- predict(lasso_final, as.matrix(X_train_scaled))
lasso_test_pred <- predict(lasso_final, as.matrix(X_test_scaled))

# Compute MSE for Lasso Regression
lasso_train_mse <- mean((lasso_train_pred - y_train)^2)
lasso_test_mse <- mean((lasso_test_pred - y_test)^2)

# Print results
cat("Ridge Regression - Training MSE:", ridge_train_mse, "\n")
```

```
## Ridge Regression - Training MSE: 1591.051
```

```r
cat("Ridge Regression - Test MSE:", ridge_test_mse, "\n")
```

```
## Ridge Regression - Test MSE: NaN
```

```r
cat("Lasso Regression - Training MSE:", lasso_train_mse, "\n")
```

```
## Lasso Regression - Training MSE: 1574.805
```

```r
cat("Lasso Regression - Test MSE:", lasso_test_mse, "\n")
```

```
## Lasso Regression - Test MSE: NaN
```

```r
# data prep
df_numeric <- data_cleaned %>% select(where(is.numeric))
# CORRELATION ANALYSIS
# Compute the correlation matrix
cor_matrix <- cor(df_numeric, use = "pairwise.complete.obs") # View Heatmap
print(cor_matrix)
```

```
##                          intersects_msa cur_smoke_q1 cur_smoke_q2  cur_smoke_q3
## intersects_msa              1.000000000  0.227580767  0.236271092  0.1983036563
## cur_smoke_q1                0.227580767  1.000000000  0.589462529  0.4403108584
## cur_smoke_q2                0.236271092  0.589462529  1.000000000  0.4499252832
## cur_smoke_q3                0.198303656  0.440310858  0.449925283  1.0000000000
## cur_smoke_q4                0.184636127  0.393426770  0.371030895  0.2915942902
## bmi_obese_q1                0.231751997  0.589120939  0.556553506  0.4415520314
## bmi_obese_q2                0.211136055  0.596645491  0.539931037  0.4374416140
## bmi_obese_q3                0.183099633  0.530408408  0.464548530  0.3501696275
## bmi_obese_q4                0.159982696  0.421801822  0.432944110  0.3125840082
## exercise_any_q1             0.291305815  0.642668289  0.592347833  0.4371979950
## exercise_any_q2             0.311602968  0.723654445  0.620150863  0.4525226048
## exercise_any_q3             0.312661164  0.687387973  0.645153280  0.4668116562
## exercise_any_q4             0.301175928  0.696704422  0.624659738  0.4759932175
## brfss_mia                  -0.310071025 -0.735159552 -0.665223506 -0.5242410817
## puninsured2010             -0.136028716 -0.128211857 -0.142835119 -0.0985862566
## reimb_penroll_adj10         0.086817246  0.040588540  0.041937951  0.0640758177
## mort_30day_hosp_z          -0.119696311 -0.038523858 -0.044894319 -0.0210292977
## adjmortmeas_amiall30day    -0.118316394 -0.063932827 -0.080695970 -0.0426095238
## adjmortmeas_chfall30day    -0.097153481 -0.032285031 -0.038472649 -0.0532964343
## med_prev_qual_z             0.177611028  0.043748263  0.061301425  0.0263286616
## primcarevis_10             -0.017707691  0.001670771 -0.005508868  0.0071302790
## diab_hemotest_10            0.078951840  0.013666312  0.021733264  0.0214808017
## diab_eyeexam_10             0.010184976 -0.051486428 -0.028568768 -0.0486912908
## diab_lipids_10              0.223426157  0.084222206  0.097029930  0.0755424181
## mammogram_10                0.158008815  0.083876376  0.102930093  0.0379305427
```

10

```
## cs00_seg_inc              0.372506982  0.258390881  0.232344584  0.1820504397
## cs00_seg_inc_pov25        0.363633017  0.260165183  0.234508205  0.1827645696
## cs00_seg_inc_aff75        0.360838434  0.241660871  0.217167390  0.1724992203
## cs_race_theil_2000        0.206448917  0.205957802  0.191532781  0.1771477287
## gini99                    0.089415925  0.151256417  0.131865909  0.0977955262
## poor_share               -0.231042266 -0.101526836 -0.106168822 -0.0428505104
## inc_share_1perc           0.086665770  0.078453659  0.083059008  0.0217312006
## frac_middleclass         -0.111170621 -0.063821871 -0.076589229 -0.0797261907
## scap_ski90pcm            -0.243785727 -0.180882543 -0.156767374 -0.1443600447
## rel_tot                  -0.213466515 -0.210979529 -0.184060194 -0.1511788050
## cs_frac_black             0.090593534 -0.020599038  0.016212725  0.0705830554
## cs_frac_hisp             -0.004006703 -0.125285439 -0.113503466 -0.0838006799
## unemp_rate               -0.072399503  0.022661893  0.004925945  0.0244893170
## cs_labforce               0.232180037  0.131421941  0.121716641  0.0708722900
## cs_elf_ind_man            0.115605221  0.129822353  0.130468038  0.1258090934
## cs_born_foreign           0.138178472 -0.017175798 -0.012460400 -0.0003196703
## mig_inflow                0.452631709  0.225342025  0.228554675  0.1766008931
## mig_outflow               0.440223440  0.195595346  0.179859803  0.1440594992
## pop_density               0.092650187  0.029535605  0.040199296  0.0377317694
## frac_traveltime_lt15     -0.493433447 -0.214749422 -0.244463248 -0.1983879240
## hhinc00                   0.257441601  0.074130635  0.091876319  0.0384608081
## median_house_value        0.281379447  0.149351428  0.158881397  0.0894930255
## ccd_exp_tot              -0.010579817 -0.102852758 -0.082907176 -0.0823934742
## score_r                  -0.135150570 -0.070624844 -0.082437064 -0.0824476086
## cs_fam_wkidsinglemom      0.062753741  0.115573074  0.110207326  0.1261769756
## subcty_exp_pc            -0.060726019 -0.070138557 -0.043861509 -0.0379915865
## taxrate                  -0.095147377 -0.161371988 -0.150469778 -0.1073668157
## tax_st_diff_top20        -0.005187465 -0.051698663 -0.029537063 -0.0330790939
## deathspc                  0.138028411  0.035314904  0.050725931  0.0709302683
##                         cur_smoke_q4 bmi_obese_q1  bmi_obese_q2 bmi_obese_q3
## intersects_msa            0.184636127  0.231751997  0.2111360550  0.183099633
## cur_smoke_q1              0.393426770  0.589120939  0.5966454908  0.530408408
## cur_smoke_q2              0.371030895  0.556553506  0.5399310370  0.464548530
## cur_smoke_q3              0.291594290  0.441552031  0.4374416140  0.350169628
## cur_smoke_q4              1.000000000  0.405154916  0.3711804966  0.291928072
## bmi_obese_q1              0.405154916  1.000000000  0.6644220992  0.547504596
## bmi_obese_q2              0.371180497  0.664422099  1.0000000000  0.545802971
## bmi_obese_q3              0.291928072  0.547504596  0.5458029714  1.000000000
## bmi_obese_q4              0.227700276  0.502409485  0.4744207691  0.402888887
## exercise_any_q1           0.384015433  0.639905182  0.6477165706  0.539969488
## exercise_any_q2           0.413532476  0.702543737  0.6818700455  0.573139694
## exercise_any_q3           0.414133066  0.677435363  0.6602007689  0.553370590
## exercise_any_q4           0.398190012  0.671762405  0.6609682171  0.545600614
## brfss_mia                -0.469824582 -0.726899201 -0.7132690824 -0.606992535
## puninsured2010           -0.053752019 -0.102916468 -0.0867959866 -0.088370816
## reimb_penroll_adj10       0.083060728  0.065899940  0.0588245245  0.024832055
## mort_30day_hosp_z        -0.034849618 -0.002203827  0.0133801736  0.024391538
## adjmortmeas_amiall30day -0.050560768 -0.044050828 -0.0253338679 -0.012802063
## adjmortmeas_chfall30day -0.025354082 -0.038694721 -0.0182535151 -0.013940869
## med_prev_qual_z           0.022825543  0.069247067  0.0538948278  0.053262282
## primcarevis_10            0.009667215  0.043625198  0.0321697821  0.022365696
## diab_hemotest_10          0.021610906  0.042695210  0.0354059621  0.042503748
## diab_eyeexam_10          -0.034604027 -0.043007483 -0.0441445493 -0.033769342
## diab_lipids_10            0.066970829  0.121443268  0.1080588341  0.086102973
```

```
## mammogram_10            0.032952600  0.085773303  0.0533068895  0.080223246
## cs00_seg_inc            0.124902789  0.183402321  0.1604267795  0.120921193
## cs00_seg_inc_pov25      0.125636463  0.189003550  0.1682251663  0.125732733
## cs00_seg_inc_aff75      0.115283745  0.168374492  0.1466606530  0.109452496
## cs_race_theil_2000      0.154328790  0.212436900  0.1820333863  0.140159759
## gini99                  0.068971573  0.158904358  0.1144192703  0.084137863
## poor_share             -0.064623294 -0.017151073 -0.0266873388 -0.036470389
## inc_share_1perc         0.025420042  0.056459940  0.0223518046  0.001971636
## frac_middleclass       -0.036211712 -0.107580542 -0.0666519145 -0.046696143
## scap_ski90pcm          -0.150803666 -0.201641715 -0.1788139756 -0.160508996
## rel_tot                -0.104058211 -0.177783551 -0.1612775223 -0.132825619
## cs_frac_black           0.027846374  0.138863109  0.0952505569  0.067887235
## cs_frac_hisp           -0.070966775 -0.068376538 -0.0605946989 -0.066306612
## unemp_rate              0.018576168  0.092100467  0.0899014279  0.076033746
## cs_labforce             0.063153298  0.039697172  0.0513249509  0.061031687
## cs_elf_ind_man          0.098057828  0.203881212  0.1928728148  0.188846169
## cs_born_foreign        -0.010941063 -0.004600049 -0.0033034345 -0.021678132
## mig_inflow              0.164952591  0.169544293  0.1779591350  0.136858480
## mig_outflow             0.130088760  0.133855772  0.1373385935  0.111609914
## pop_density             0.027634317  0.012955102 -0.0002068472 -0.003737283
## frac_traveltime_lt15   -0.181182122 -0.214271697 -0.1987472180 -0.157411448
## hhinc00                 0.037588560 -0.015309513 -0.0228336766 -0.014294041
## median_house_value      0.077180115  0.088618795  0.0640117679  0.049756874
## ccd_exp_tot            -0.091511257 -0.131274569 -0.1400227311 -0.121272646
## score_r                -0.069796803 -0.108455207 -0.0823851842 -0.059011185
## cs_fam_wkidsinglemom    0.070502633  0.191762489  0.1482968450  0.112730659
## subcty_exp_pc          -0.070768586 -0.065287052 -0.0768335160 -0.063705498
## taxrate                -0.125859404 -0.175339820 -0.1759949822 -0.152480688
## tax_st_diff_top20      -0.031750077 -0.004540564  0.0029729179 -0.024956112
## deathspc                0.040896587  0.063689015  0.0464543106  0.001830396
##                        bmi_obese_q4 exercise_any_q1 exercise_any_q2
## intersects_msa          0.159982696      0.29130581      0.31160297
## cur_smoke_q1            0.421801822      0.64266829      0.72365445
## cur_smoke_q2            0.432944110      0.59234783      0.62015086
## cur_smoke_q3            0.312584008      0.43719799      0.45252260
## cur_smoke_q4            0.227700276      0.38401543      0.41353248
## bmi_obese_q1            0.502409485      0.63990518      0.70254374
## bmi_obese_q2            0.474420769      0.64771657      0.68187005
## bmi_obese_q3            0.402888887      0.53996949      0.57313969
## bmi_obese_q4            1.000000000      0.48079208      0.53502507
## exercise_any_q1         0.480792076      1.00000000      0.88018929
## exercise_any_q2         0.535025072      0.88018929      1.00000000
## exercise_any_q3         0.516272658      0.86652815      0.88715518
## exercise_any_q4         0.513647975      0.85240212      0.88142077
## brfss_mia              -0.572493062     -0.86379282     -0.90977734
## puninsured2010         -0.052619656     -0.18996275     -0.18835953
## reimb_penroll_adj10     0.047833792     -0.15156789     -0.12023527
## mort_30day_hosp_z       0.016847135     -0.09133451     -0.08280771
## adjmortmeas_amiall30day -0.007821655     -0.13903060     -0.12552175
## adjmortmeas_chfall30day -0.001920917     -0.01382799     -0.02266073
## med_prev_qual_z         0.006130325      0.19849843      0.18598982
## primcarevis_10         -0.006350504     -0.07047574     -0.06319572
## diab_hemotest_10        0.018726580      0.07382662      0.07844232
## diab_eyeexam_10        -0.055270423      0.10347753      0.08600477
```

```
## diab_lipids_10           0.049392108      0.11943933      0.13114178
## mammogram_10             0.015241171      0.22421015      0.21694490
## cs00_seg_inc             0.086968957      0.36089385      0.36508903
## cs00_seg_inc_pov25       0.097033303      0.35577903      0.36054159
## cs00_seg_inc_aff75       0.073468948      0.34887481      0.35195041
## cs_race_theil_2000       0.109612826      0.22971241      0.24858847
## gini99                   0.065333007      0.15393334      0.17266619
## poor_share              -0.025622937     -0.20125901     -0.18769832
## inc_share_1perc          0.013732505      0.15294996      0.15053421
## frac_middleclass        -0.028961565     -0.05836083     -0.06954240
## scap_ski90pcm           -0.154668322     -0.09866448     -0.13607422
## rel_tot                 -0.109412704     -0.21201245     -0.21450878
## cs_frac_black            0.024931939     -0.02760198     -0.01165834
## cs_frac_hisp            -0.042148326     -0.04467264     -0.05317880
## unemp_rate               0.062768680     -0.03896177     -0.01653932
## cs_labforce              0.038663538      0.27214123      0.25627569
## cs_elf_ind_man           0.146808601      0.05500369      0.09056310
## cs_born_foreign         -0.011187470      0.13564740      0.12951706
## mig_inflow               0.130027981      0.33421060      0.33202357
## mig_outflow              0.101085351      0.30067729      0.30390842
## pop_density             -0.011881473      0.06193277      0.06253166
## frac_traveltime_lt15    -0.126294346     -0.16524592     -0.19836298
## hhinc00                 -0.034092994      0.22421890      0.20449362
## median_house_value       0.035258512      0.34849762      0.32817593
## ccd_exp_tot             -0.115148918     -0.05244010     -0.06140959
## score_r                 -0.033878814     -0.06071427     -0.07375416
## cs_fam_wkidsinglemom     0.071447556      0.08348039      0.09593891
## subcty_exp_pc           -0.057609858      0.02860255      0.01506953
## taxrate                 -0.136384966     -0.09751151     -0.11720918
## tax_st_diff_top20        0.004805917      0.03078945      0.02103331
## deathspc                 0.016635032      0.05823319      0.07801299
##                         exercise_any_q3 exercise_any_q4    brfss_mia
## intersects_msa               0.31266116     0.301175928 -0.310071025
## cur_smoke_q1                 0.68738797     0.696704422 -0.735159552
## cur_smoke_q2                 0.64515328     0.624659738 -0.665223506
## cur_smoke_q3                 0.46681166     0.475993218 -0.524241082
## cur_smoke_q4                 0.41413307     0.398190012 -0.469824582
## bmi_obese_q1                 0.67743536     0.671762405 -0.726899201
## bmi_obese_q2                 0.66020077     0.660968217 -0.713269082
## bmi_obese_q3                 0.55337059     0.545600614 -0.606992535
## bmi_obese_q4                 0.51627266     0.513647975 -0.572493062
## exercise_any_q1              0.86652815     0.852402118 -0.863792822
## exercise_any_q2              0.88715518     0.881420772 -0.909777335
## exercise_any_q3              1.00000000     0.874856082 -0.916738758
## exercise_any_q4              0.87485608     1.000000000 -0.931850341
## brfss_mia                   -0.91673876    -0.931850341  1.000000000
## puninsured2010              -0.19963724    -0.201400158  0.187118221
## reimb_penroll_adj10         -0.11493661    -0.097997629  0.060412725
## mort_30day_hosp_z           -0.07384760    -0.091795045  0.061521652
## adjmortmeas_amiall30day     -0.12132053    -0.134511475  0.111498154
## adjmortmeas_chfall30day     -0.01104142    -0.029004385  0.015620917
## med_prev_qual_z              0.19004803     0.192770024 -0.158922579
## primcarevis_10              -0.06174192    -0.061671313  0.052413019
## diab_hemotest_10             0.08385792     0.084101123 -0.070525590
```

```
## diab_eyeexam_10            0.08696669     0.090580170 -0.053276441
## diab_lipids_10             0.13445994     0.150893888 -0.139284080
## mammogram_10               0.22093204     0.225139025 -0.190076011
## cs00_seg_inc               0.36535515     0.375899567 -0.359338428
## cs00_seg_inc_pov25         0.35880464     0.368748025 -0.355236126
## cs00_seg_inc_aff75         0.35496412     0.364509566 -0.345928657
## cs_race_theil_2000         0.25643207     0.262476074 -0.271653041
## gini99                     0.16393313     0.189124885 -0.183790879
## poor_share                -0.19534442    -0.200811654  0.170945913
## inc_share_1perc            0.14763980     0.162157455 -0.136470698
## frac_middleclass          -0.06623222    -0.081381072  0.086510117
## scap_ski90pcm             -0.12215812    -0.132847369  0.161603417
## rel_tot                   -0.21392552    -0.213200156  0.208146784
## cs_frac_black             -0.01039616    -0.007822013 -0.011990901
## cs_frac_hisp              -0.06763066    -0.057904065  0.071554380
## unemp_rate                -0.02459837    -0.034595498  0.008929087
## cs_labforce                0.25070318     0.260842068 -0.234538464
## cs_elf_ind_man             0.09742190     0.096968094 -0.121843326
## cs_born_foreign            0.11316891     0.129285590 -0.107223225
## mig_inflow                 0.31942720     0.326737938 -0.319990830
## mig_outflow                0.28212217     0.286672417 -0.280927547
## pop_density                0.05955735     0.065484703 -0.062299995
## frac_traveltime_lt15      -0.19922850    -0.209953342  0.222688379
## hhinc00                    0.20717779     0.227769740 -0.181984191
## median_house_value         0.32507797     0.335714908 -0.301011917
## ccd_exp_tot               -0.05527474    -0.050267103  0.082339643
## score_r                   -0.07374422    -0.076834165  0.094659005
## cs_fam_wkidsinglemom       0.09218441     0.099638139 -0.115091618
## subcty_exp_pc              0.01565130     0.024200065  0.010208698
## taxrate                   -0.11855593    -0.104200914  0.144250864
## tax_st_diff_top20          0.02721704     0.025087986 -0.020460431
## deathspc                   0.06604410     0.072588816 -0.080493124
##                         puninsured2010 reimb_penroll_adj10 mort_30day_hosp_z
## intersects_msa            -0.136028716         0.086817246      -0.119696311
## cur_smoke_q1              -0.128211857         0.040588540      -0.038523858
## cur_smoke_q2              -0.142835119         0.041937951      -0.044894319
## cur_smoke_q3              -0.098586257         0.064075818      -0.021029298
## cur_smoke_q4              -0.053752019         0.083060728      -0.034849618
## bmi_obese_q1              -0.102916468         0.065899940      -0.002203827
## bmi_obese_q2              -0.086795987         0.058824525       0.013380174
## bmi_obese_q3              -0.088370816         0.024832055       0.024391538
## bmi_obese_q4              -0.052619656         0.047833792       0.016847135
## exercise_any_q1           -0.189962749        -0.151567887      -0.091334506
## exercise_any_q2           -0.188359526        -0.120235267      -0.082807708
## exercise_any_q3           -0.199637238        -0.114936610      -0.073847595
## exercise_any_q4           -0.201400158        -0.097997629      -0.091795045
## brfss_mia                  0.187118221         0.060412725       0.061521652
## puninsured2010             1.000000000         0.336524422       0.197722167
## reimb_penroll_adj10        0.336524422         1.000000000       0.003210837
## mort_30day_hosp_z          0.197722167         0.003210837       1.000000000
## adjmortmeas_amiall30day    0.172072655         0.109683242       0.732008041
## adjmortmeas_chfall30day    0.087016568        -0.187980260       0.670198591
## med_prev_qual_z           -0.445215939        -0.416111485      -0.168616556
## primcarevis_10             0.028152078         0.138434964       0.053589075
```

```
## diab_hemotest_10          -0.352097493         -0.186688531        -0.106444367
## diab_eyeexam_10           -0.398221861         -0.397743325        -0.142342551
## diab_lipids_10            -0.219501355          0.010018568        -0.127965090
## mammogram_10              -0.484134294         -0.369693970        -0.215833908
## cs00_seg_inc              -0.174163837         -0.007014490        -0.180264725
## cs00_seg_inc_pov25        -0.185818290         -0.012585748        -0.167860484
## cs00_seg_inc_aff75        -0.157437906         -0.007902860        -0.184370779
## cs_race_theil_2000         0.007343001          0.078916218        -0.092914968
## gini99                     0.343077188          0.232032725         0.035194124
## poor_share                 0.540947425          0.348280509         0.210061801
## inc_share_1perc            0.008714516          0.009721595        -0.101293263
## frac_middleclass          -0.359879680         -0.328644206        -0.053945177
## scap_ski90pcm             -0.483527964         -0.387633768        -0.147645261
## rel_tot                   -0.011993080          0.032099397        -0.003365403
## cs_frac_black              0.193781721          0.216181860         0.111778203
## cs_frac_hisp               0.463160367          0.083394847         0.014754820
## unemp_rate                 0.287841413          0.189100589         0.156051745
## cs_labforce               -0.449901540         -0.345372203        -0.198792693
## cs_elf_ind_man            -0.158570589          0.097768330         0.114825511
## cs_born_foreign            0.277441636         -0.022579310        -0.072037170
## mig_inflow                -0.098503059          0.003512911        -0.133248188
## mig_outflow               -0.055554514          0.001013416        -0.122422455
## pop_density               -0.058207239          0.016093664        -0.092649801
## frac_traveltime_lt15       0.056300871         -0.265178830         0.092853893
## hhinc00                   -0.423360151         -0.266352932        -0.276139282
## median_house_value        -0.241131707         -0.238039215        -0.217271238
## ccd_exp_tot               -0.159423970         -0.182632484        -0.127295040
## score_r                   -0.230729171         -0.150312298        -0.075175362
## cs_fam_wkidsinglemom       0.199470148          0.171407957         0.113858760
## subcty_exp_pc             -0.060613384         -0.216208754        -0.076202926
## taxrate                    0.056256946         -0.115412915        -0.096191238
## tax_st_diff_top20         -0.229539579         -0.137556826        -0.036707309
## deathspc                  -0.044149440          0.076199776        -0.016912467
##                          adjmortmeas_amiall30day adjmortmeas_chfall30day
## intersects_msa                      -0.118316394            -0.097153481
## cur_smoke_q1                        -0.063932827            -0.032285031
## cur_smoke_q2                        -0.080695970            -0.038472649
## cur_smoke_q3                        -0.042609524            -0.053296434
## cur_smoke_q4                        -0.050560768            -0.025354082
## bmi_obese_q1                        -0.044050828            -0.038694721
## bmi_obese_q2                        -0.025333868            -0.018253515
## bmi_obese_q3                        -0.012802063            -0.013940869
## bmi_obese_q4                        -0.007821655            -0.001920917
## exercise_any_q1                     -0.139030602            -0.013827988
## exercise_any_q2                     -0.125521753            -0.022660733
## exercise_any_q3                     -0.121320531            -0.011041422
## exercise_any_q4                     -0.134511475            -0.029004385
## brfss_mia                            0.111498154             0.015620917
## puninsured2010                       0.172072655             0.087016568
## reimb_penroll_adj10                  0.109683242            -0.187980260
## mort_30day_hosp_z                    0.732008041             0.670198591
## adjmortmeas_amiall30day              1.000000000             0.218354347
## adjmortmeas_chfall30day              0.218354347             1.000000000
## med_prev_qual_z                     -0.194238412            -0.051311798
```

```
## primcarevis_10                 0.033840417            -0.013130615
## diab_hemotest_10              -0.111478857            -0.053173746
## diab_eyeexam_10               -0.133848840            -0.004634422
## diab_lipids_10                -0.117123916            -0.135344176
## mammogram_10                  -0.211626559            -0.087314380
## cs00_seg_inc                  -0.169518036            -0.141001326
## cs00_seg_inc_pov25            -0.159463412            -0.132982869
## cs00_seg_inc_aff75            -0.175402240            -0.138701275
## cs_race_theil_2000            -0.077266445            -0.113359501
## gini99                         0.010182423            -0.087564361
## poor_share                     0.195557435             0.012735173
## inc_share_1perc               -0.089833608            -0.081991205
## frac_middleclass              -0.034591709             0.117049007
## scap_ski90pcm                 -0.111692194             0.058230569
## rel_tot                        0.061315922             0.060494388
## cs_frac_black                  0.073087239            -0.111797878
## cs_frac_hisp                   0.010058501             0.007898751
## unemp_rate                     0.131556918             0.000511416
## cs_labforce                   -0.201237907             0.005905577
## cs_elf_ind_man                 0.080896479             0.033421411
## cs_born_foreign               -0.067546999            -0.043056862
## mig_inflow                    -0.142037358            -0.088842802
## mig_outflow                   -0.125353247            -0.092371179
## pop_density                   -0.068447596            -0.096123827
## frac_traveltime_lt15           0.087424673             0.190753764
## hhinc00                       -0.246590799            -0.107392331
## median_house_value            -0.223939328            -0.088972651
## ccd_exp_tot                   -0.109129962            -0.039079969
## score_r                       -0.030744716             0.059591695
## cs_fam_wkidsinglemom           0.061563075            -0.079196110
## subcty_exp_pc                 -0.086148806             0.007297579
## taxrate                       -0.066716196            -0.031926399
## tax_st_diff_top20             -0.045117875             0.009887240
## deathspc                      -0.019227729            -0.071569242
##                         med_prev_qual_z primcarevis_10 diab_hemotest_10
## intersects_msa              0.177611028  -0.0177076913      0.078951840
## cur_smoke_q1                0.043748263   0.0016707714      0.013666312
## cur_smoke_q2                0.061301425  -0.0055088684      0.021733264
## cur_smoke_q3                0.026328662   0.0071302790      0.021480802
## cur_smoke_q4                0.022825543   0.0096672153      0.021610906
## bmi_obese_q1                0.069247067   0.0436251978      0.042695210
## bmi_obese_q2                0.053894828   0.0321697821      0.035405962
## bmi_obese_q3                0.053262282   0.0223656961      0.042503748
## bmi_obese_q4                0.006130325  -0.0063505042      0.018726580
## exercise_any_q1             0.198498431  -0.0704757449      0.073826618
## exercise_any_q2             0.185989824  -0.0631957248      0.078442322
## exercise_any_q3             0.190048029  -0.0617419161      0.083857920
## exercise_any_q4             0.192770024  -0.0616713134      0.084101123
## brfss_mia                  -0.158922579   0.0524130192     -0.070525590
## puninsured2010             -0.445215939   0.0281520784     -0.352097493
## reimb_penroll_adj10        -0.416111485   0.1384349644     -0.186688531
## mort_30day_hosp_z          -0.168616556   0.0535890749     -0.106444367
## adjmortmeas_amiall30day    -0.194238412   0.0338404168     -0.111478857
## adjmortmeas_chfall30day    -0.051311798  -0.0131306150     -0.053173746
```

```
## med_prev_qual_z             1.000000000  0.2806218805      0.749479191
## primcarevis_10              0.280621881  1.0000000000      0.134382777
## diab_hemotest_10            0.749479191  0.1343827775      1.000000000
## diab_eyeexam_10             0.664944562  0.0436036410      0.327232103
## diab_lipids_10              0.658606290  0.1043710959      0.654737342
## mammogram_10                0.758386227  0.0348100631      0.457211138
## cs00_seg_inc                0.191370232 -0.0758262205      0.021354861
## cs00_seg_inc_pov25          0.187210928 -0.0602724081      0.024607830
## cs00_seg_inc_aff75          0.193379260 -0.0865637377      0.024944190
## cs_race_theil_2000         -0.034114305 -0.0404094268     -0.140331421
## gini99                     -0.100036036  0.0137370760     -0.137710631
## poor_share                 -0.499625321  0.0866434771     -0.350231000
## inc_share_1perc             0.087424001 -0.0492597441      0.006147421
## frac_middleclass            0.205041401 -0.0607020882      0.207265114
## scap_ski90pcm               0.300099321 -0.0667434476      0.230755646
## rel_tot                     0.001384724  0.0001122557     -0.010250902
## cs_frac_black              -0.078954814  0.1492721042     -0.090948240
## cs_frac_hisp               -0.168967851 -0.1073772713     -0.181599020
## unemp_rate                 -0.280774695  0.0289989311     -0.156611855
## cs_labforce                 0.388507471 -0.0928801518      0.221667037
## cs_elf_ind_man              0.074016169  0.1274769690      0.178633609
## cs_born_foreign             0.034404730 -0.1719785952     -0.044242180
## mig_inflow                  0.189602583 -0.0342241734      0.059695410
## mig_outflow                 0.119758304 -0.0629830259      0.001239106
## pop_density                 0.014890153 -0.1131919475     -0.005374482
## frac_traveltime_lt15       -0.058831916 -0.0609765492     -0.082109742
## hhinc00                     0.463089027 -0.1088606725      0.261600664
## median_house_value          0.265366704 -0.1774164394      0.100942333
## ccd_exp_tot                 0.117845901 -0.1027559470      0.057855931
## score_r                     0.170418390 -0.0225766254      0.177402963
## cs_fam_wkidsinglemom       -0.164008443  0.0730926671     -0.186427359
## subcty_exp_pc               0.052046373 -0.1758260370     -0.010876380
## taxrate                     0.060840183 -0.0913662356      0.008928511
## tax_st_diff_top20           0.088846217 -0.0970312659      0.037649435
## deathspc                   -0.014615348 -0.0428328977     -0.076765746
##                          diab_eyeexam_10 diab_lipids_10 mammogram_10
## intersects_msa              0.010184976  0.2234261567  0.158008815
## cur_smoke_q1               -0.051486428  0.0842222064  0.083876376
## cur_smoke_q2               -0.028568768  0.0970299296  0.102930093
## cur_smoke_q3               -0.048691291  0.0755424181  0.037930543
## cur_smoke_q4               -0.034604027  0.0669708291  0.032952600
## bmi_obese_q1               -0.043007483  0.1214432681  0.085773303
## bmi_obese_q2               -0.044144549  0.1080588341  0.053306890
## bmi_obese_q3               -0.033769342  0.0861029735  0.080223246
## bmi_obese_q4               -0.055270423  0.0493921084  0.015241171
## exercise_any_q1             0.103477528  0.1194393304  0.224210148
## exercise_any_q2             0.086004773  0.1311417766  0.216944896
## exercise_any_q3             0.086966687  0.1344599440  0.220932045
## exercise_any_q4             0.090580170  0.1508938878  0.225139025
## brfss_mia                  -0.053276441 -0.1392840803 -0.190076011
## puninsured2010             -0.398221861 -0.2195013548 -0.484134294
## reimb_penroll_adj10        -0.397743325  0.0100185677 -0.369693970
## mort_30day_hosp_z          -0.142342551 -0.1279650904 -0.215833908
## adjmortmeas_amiall30day    -0.133848840 -0.1171239164 -0.211626559
```

```
## adjmortmeas_chfall30day    -0.004634422  -0.1353441764 -0.087314380
## med_prev_qual_z             0.664944562   0.6586062904  0.758386227
## primcarevis_10              0.043603641   0.1043710959  0.034810063
## diab_hemotest_10            0.327232103   0.6547373420  0.457211138
## diab_eyeexam_10             1.000000000   0.1675734815  0.526199585
## diab_lipids_10              0.167573482   1.0000000000  0.371017206
## mammogram_10                0.526199585   0.3710172057  1.000000000
## cs00_seg_inc                0.131726291   0.1230251705  0.221451097
## cs00_seg_inc_pov25          0.129868499   0.1149592390  0.216918615
## cs00_seg_inc_aff75          0.130982390   0.1280782084  0.219499286
## cs_race_theil_2000         -0.016325123  -0.0535616590  0.053788536
## gini99                     -0.164195880   0.0405097560 -0.059520334
## poor_share                 -0.415594277  -0.2281144587 -0.476709905
## inc_share_1perc             0.034892883   0.0885521700  0.107406856
## frac_middleclass            0.262012704  -0.0178519600  0.199987671
## scap_ski90pcm               0.415432760  -0.0332724864  0.323055675
## rel_tot                     0.155902390  -0.0845562341 -0.008672202
## cs_frac_black              -0.132014378  -0.0151748242 -0.075761844
## cs_frac_hisp               -0.130230010  -0.0382472128 -0.252186967
## unemp_rate                 -0.324861822  -0.0394663235 -0.266605405
## cs_labforce                 0.372360764   0.0950422985  0.365582667
## cs_elf_ind_man             -0.068316346   0.1743179082  0.033905420
## cs_born_foreign             0.030116800   0.0766547419 -0.013727364
## mig_inflow                  0.053306161   0.1992174631  0.152695661
## mig_outflow                 0.027634786   0.1199026008  0.079097751
## pop_density                 0.034201299   0.0379318265  0.026320736
## frac_traveltime_lt15        0.186136341  -0.2958436710 -0.027252249
## hhinc00                     0.383007535   0.2375166539  0.466815553
## median_house_value          0.156691819   0.1663808047  0.294818503
## ccd_exp_tot                 0.156136461   0.0051544002  0.151471867
## score_r                     0.236246956   0.0377458035  0.152556306
## cs_fam_wkidsinglemom       -0.181105404  -0.0703520906 -0.115130276
## subcty_exp_pc               0.097899410  -0.0861387842  0.112163964
## taxrate                     0.101894008  -0.0633819350  0.075261910
## tax_st_diff_top20           0.138399151   0.0321530504  0.121740299
## deathspc                   -0.004647352  -0.0001429391  0.031810846
##                          cs00_seg_inc cs00_seg_inc_pov25 cs00_seg_inc_aff75
## intersects_msa             0.37250698         0.36363302         0.36083843
## cur_smoke_q1               0.25839088         0.26016518         0.24166087
## cur_smoke_q2               0.23234458         0.23450820         0.21716739
## cur_smoke_q3               0.18205044         0.18276457         0.17249922
## cur_smoke_q4               0.12490279         0.12563646         0.11528375
## bmi_obese_q1               0.18340232         0.18900355         0.16837449
## bmi_obese_q2               0.16042678         0.16822517         0.14666065
## bmi_obese_q3               0.12092119         0.12573273         0.10945250
## bmi_obese_q4               0.08696896         0.09703330         0.07346895
## exercise_any_q1            0.36089385         0.35577903         0.34887481
## exercise_any_q2            0.36508903         0.36054159         0.35195041
## exercise_any_q3            0.36535515         0.35880464         0.35496412
## exercise_any_q4            0.37589957         0.36874802         0.36450957
## brfss_mia                 -0.35933843        -0.35523613        -0.34592866
## puninsured2010            -0.17416384        -0.18581829        -0.15743791
## reimb_penroll_adj10       -0.00701449        -0.01258575        -0.00790286
## mort_30day_hosp_z         -0.18026473        -0.16786048        -0.18437078
```

```
## adjmortmeas_amiall30day  -0.16951804        -0.15946341        -0.17540224
## adjmortmeas_chfall30day  -0.14100133        -0.13298287        -0.13870127
## med_prev_qual_z           0.19137023         0.18721093         0.19337926
## primcarevis_10           -0.07582622        -0.06027241        -0.08656374
## diab_hemotest_10          0.02135486         0.02460783         0.02494419
## diab_eyeexam_10           0.13172629         0.12986850         0.13098239
## diab_lipids_10            0.12302517         0.11495924         0.12807821
## mammogram_10              0.22145110         0.21691861         0.21949929
## cs00_seg_inc              1.00000000         0.97326489         0.97961259
## cs00_seg_inc_pov25        0.97326489         1.00000000         0.91486274
## cs00_seg_inc_aff75        0.97961259         0.91486274         1.00000000
## cs_race_theil_2000        0.55919392         0.55576042         0.52950234
## gini99                    0.39745993         0.36603218         0.40253197
## poor_share               -0.14743927        -0.12522384        -0.16628935
## inc_share_1perc           0.36796766         0.32615931         0.38396694
## frac_middleclass         -0.32265534        -0.29294510        -0.32594605
## scap_ski90pcm            -0.08945468        -0.08762761        -0.08069814
## rel_tot                  -0.13352620        -0.14232381        -0.11724724
## cs_frac_black             0.12990330         0.13778973         0.10964536
## cs_frac_hisp              0.12486947         0.09330163         0.15227935
## unemp_rate               -0.14669794        -0.13573792        -0.15275655
## cs_labforce               0.33710458         0.32640731         0.33649679
## cs_elf_ind_man           -0.15123452        -0.14213576        -0.15857893
## cs_born_foreign           0.39748972         0.33906916         0.42640723
## mig_inflow                0.38440255         0.36133275         0.38863482
## mig_outflow               0.45209173         0.42812012         0.45358067
## pop_density               0.22630318         0.20732714         0.22734814
## frac_traveltime_lt15     -0.23614877        -0.21578669        -0.23250251
## hhinc00                   0.43551555         0.37361098         0.47209249
## median_house_value        0.42676101         0.38525695         0.44120787
## ccd_exp_tot               0.05537670         0.05293956         0.05523500
## score_r                  -0.20590096        -0.18099643        -0.21020634
## cs_fam_wkidsinglemom      0.24938585         0.26271808         0.21993987
## subcty_exp_pc             0.16744698         0.15600652         0.17356070
## taxrate                   0.04383120         0.03988167         0.05332386
## tax_st_diff_top20         0.05762270         0.04455916         0.06337997
## deathspc                  0.21171087         0.18898177         0.21094128
##                         cs_race_theil_2000        gini99  poor_share
## intersects_msa                 0.206448917  0.0894159252 -0.23104227
## cur_smoke_q1                   0.205957802  0.1512564169 -0.10152684
## cur_smoke_q2                   0.191532781  0.1318659085 -0.10616882
## cur_smoke_q3                   0.177147729  0.0977955262 -0.04285051
## cur_smoke_q4                   0.154328790  0.0689715731 -0.06462329
## bmi_obese_q1                   0.212436900  0.1589043584 -0.01715107
## bmi_obese_q2                   0.182033386  0.1144192703 -0.02668734
## bmi_obese_q3                   0.140159759  0.0841378633 -0.03647039
## bmi_obese_q4                   0.109612826  0.0653330075 -0.02562294
## exercise_any_q1                0.229712412  0.1539333410 -0.20125901
## exercise_any_q2                0.248588473  0.1726661858 -0.18769832
## exercise_any_q3                0.256432065  0.1639331279 -0.19534442
## exercise_any_q4                0.262476074  0.1891248855 -0.20081165
## brfss_mia                     -0.271653041 -0.1837908786  0.17094591
## puninsured2010                 0.007343001  0.3430771877  0.54094743
## reimb_penroll_adj10            0.078916218  0.2320327255  0.34828051
```

```
## mort_30day_hosp_z            -0.092914968  0.0351941243  0.21006180
## adjmortmeas_amiall30day      -0.077266445  0.0101824234  0.19555743
## adjmortmeas_chfall30day      -0.113359501 -0.0875643607  0.01273517
## med_prev_qual_z              -0.034114305 -0.1000360356 -0.49962532
## primcarevis_10               -0.040409427  0.0137370760  0.08664348
## diab_hemotest_10             -0.140331421 -0.1377106310 -0.35023100
## diab_eyeexam_10              -0.016325123 -0.1641958804 -0.41559428
## diab_lipids_10               -0.053561659  0.0405097560 -0.22811446
## mammogram_10                  0.053788536 -0.0595203343 -0.47670991
## cs00_seg_inc                  0.559193922  0.3974599336 -0.14743927
## cs00_seg_inc_pov25            0.555760419  0.3660321760 -0.12522384
## cs00_seg_inc_aff75            0.529502345  0.4025319730 -0.16628935
## cs_race_theil_2000            1.000000000  0.3615807377  0.10717571
## gini99                        0.361580738  1.0000000000  0.40132341
## poor_share                    0.107175714  0.4013234067  1.00000000
## inc_share_1perc               0.214012772  0.7060661568 -0.02490420
## frac_middleclass             -0.268565995 -0.6625682892 -0.55594447
## scap_ski90pcm                -0.116000112 -0.3779745180 -0.41922377
## rel_tot                      -0.018771685 -0.1536091791 -0.06068715
## cs_frac_black                 0.235081082  0.4518750964  0.46023925
## cs_frac_hisp                  0.088339010  0.2009538513  0.25243409
## unemp_rate                    0.075999576  0.2085590912  0.54619027
## cs_labforce                   0.012486263 -0.2362394675 -0.69180256
## cs_elf_ind_man               -0.070760246 -0.1228602188 -0.10052400
## cs_born_foreign               0.258068449  0.2960231218  0.02676097
## mig_inflow                    0.094375681  0.0773302969 -0.34907185
## mig_outflow                   0.165344028  0.1245219370 -0.24280263
## pop_density                   0.224594128  0.2388933161  0.01326314
## frac_traveltime_lt15         -0.140471968 -0.1751569529  0.06169870
## hhinc00                       0.102707316 -0.0007508972 -0.71613685
## median_house_value            0.153490315  0.1985269029 -0.38081989
## ccd_exp_tot                   0.020683675 -0.0649090972 -0.17246232
## score_r                      -0.287841385 -0.4421586040 -0.35253405
## cs_fam_wkidsinglemom          0.377836079  0.5740309146  0.58138337
## subcty_exp_pc                 0.122076479  0.0473661331 -0.13649738
## taxrate                       0.011209487 -0.0089688125 -0.12294941
## tax_st_diff_top20             0.054435915 -0.0457180324 -0.05210190
## deathspc                      0.222344332  0.2151956725  0.08079107
##                         inc_share_1perc frac_middleclass scap_ski90pcm
## intersects_msa              0.086665770      -0.11117062  -0.243785727
## cur_smoke_q1                0.078453659      -0.06382187  -0.180882543
## cur_smoke_q2                0.083059008      -0.07658923  -0.156767374
## cur_smoke_q3                0.021731201      -0.07972619  -0.144360045
## cur_smoke_q4                0.025420042      -0.03621171  -0.150803666
## bmi_obese_q1                0.056459940      -0.10758054  -0.201641715
## bmi_obese_q2                0.022351805      -0.06665191  -0.178813976
## bmi_obese_q3                0.001971636      -0.04669614  -0.160508996
## bmi_obese_q4                0.013732505      -0.02896156  -0.154668322
## exercise_any_q1             0.152949961      -0.05836083  -0.098664484
## exercise_any_q2             0.150534206      -0.06954240  -0.136074218
## exercise_any_q3             0.147639795      -0.06623222  -0.122158122
## exercise_any_q4             0.162157455      -0.08138107  -0.132847369
## brfss_mia                  -0.136470698       0.08651012   0.161603417
## puninsured2010              0.008714516      -0.35987968  -0.483527964
```

```
## reimb_penroll_adj10         0.009721595      -0.32864421  -0.387633768
## mort_30day_hosp_z          -0.101293263      -0.05394518  -0.147645261
## adjmortmeas_amiall30day    -0.089833608      -0.03459171  -0.111692194
## adjmortmeas_chfall30day    -0.081991205       0.11704901   0.058230569
## med_prev_qual_z             0.087424001       0.20504140   0.300099321
## primcarevis_10             -0.049259744      -0.06070209  -0.066743448
## diab_hemotest_10            0.006147421       0.20726511   0.230755646
## diab_eyeexam_10             0.034892883       0.26201270   0.415432760
## diab_lipids_10              0.088552170      -0.01785196  -0.033272486
## mammogram_10                0.107406856       0.19998767   0.323055675
## cs00_seg_inc                0.367967663      -0.32265534  -0.089454679
## cs00_seg_inc_pov25          0.326159308      -0.29294510  -0.087627605
## cs00_seg_inc_aff75          0.383966937      -0.32594605  -0.080698136
## cs_race_theil_2000          0.214012772      -0.26856599  -0.116000112
## gini99                      0.706066157      -0.66256829  -0.377974518
## poor_share                 -0.024904202      -0.55594447  -0.419223770
## inc_share_1perc             1.000000000      -0.29147512  -0.044054079
## frac_middleclass           -0.291475119       1.00000000   0.517355767
## scap_ski90pcm              -0.044054079       0.51735577   1.000000000
## rel_tot                    -0.047546509       0.22220071   0.407891867
## cs_frac_black               0.107960983      -0.61030662  -0.392857150
## cs_frac_hisp                0.089184328      -0.24262220  -0.210859806
## unemp_rate                 -0.040194172      -0.32484669  -0.341012495
## cs_labforce                 0.112076449       0.32637997   0.332591574
## cs_elf_ind_man             -0.074270177       0.09375352  -0.079134164
## cs_born_foreign             0.279890753      -0.26753421  -0.188552029
## mig_inflow                  0.114320728      -0.13420249  -0.300350738
## mig_outflow                 0.148069937      -0.18949747  -0.286854483
## pop_density                 0.237974313      -0.12195298  -0.017206695
## frac_traveltime_lt15       -0.081784149       0.33468762   0.489457126
## hhinc00                     0.320612239       0.05580736   0.331117525
## median_house_value          0.415356967      -0.16665800   0.008797823
## ccd_exp_tot                 0.053841388       0.06827812   0.216378601
## score_r                    -0.180321010       0.55541531   0.390679319
## cs_fam_wkidsinglemom        0.170123285      -0.63438763  -0.365209559
## subcty_exp_pc               0.149532757       0.04658251   0.198925956
## taxrate                     0.064301590       0.06221683   0.178687645
## tax_st_diff_top20           0.013880922       0.12031460   0.177558870
## deathspc                    0.150652692      -0.27628777  -0.113816339
##                               rel_tot cs_frac_black cs_frac_hisp   unemp_rate
## intersects_msa          -0.2134665151   0.090593534 -0.004006703 -0.072399503
## cur_smoke_q1            -0.2109795286  -0.020599038 -0.125285439  0.022661893
## cur_smoke_q2            -0.1840601937   0.016212725 -0.113503466  0.004925945
## cur_smoke_q3            -0.1511788050   0.070583055 -0.083800680  0.024489317
## cur_smoke_q4            -0.1040582114   0.027846374 -0.070966775  0.018576168
## bmi_obese_q1            -0.1777835507   0.138863109 -0.068376538  0.092100467
## bmi_obese_q2            -0.1612775223   0.095250557 -0.060594699  0.089901428
## bmi_obese_q3            -0.1328256195   0.067887235 -0.066306612  0.076033746
## bmi_obese_q4            -0.1094127042   0.024931939 -0.042148326  0.062768680
## exercise_any_q1         -0.2120124505  -0.027601976 -0.044672637 -0.038961765
## exercise_any_q2         -0.2145087837  -0.011658340 -0.053178795 -0.016539321
## exercise_any_q3         -0.2139255180  -0.010396160 -0.067630658 -0.024598372
## exercise_any_q4         -0.2132001563  -0.007822013 -0.057904065 -0.034595498
## brfss_mia                0.2081467844  -0.011990901  0.071554380  0.008929087
```

```
## puninsured2010          -0.0119930796  0.193781721  0.463160367  0.287841413
## reimb_penroll_adj10      0.0320993971  0.216181860  0.083394847  0.189100589
## mort_30day_hosp_z       -0.0033654027  0.111778203  0.014754820  0.156051745
## adjmortmeas_amiall30day  0.0613159217  0.073087239  0.010058501  0.131556918
## adjmortmeas_chfall30day  0.0604943878 -0.111797878  0.007898751  0.000511416
## med_prev_qual_z          0.0013847240 -0.078954814 -0.168967851 -0.280774695
## primcarevis_10           0.0001122557  0.149272104 -0.107377271  0.028998931
## diab_hemotest_10        -0.0102509017 -0.090948240 -0.181599020 -0.156611855
## diab_eyeexam_10          0.1559023901 -0.132014378 -0.130230010 -0.324861822
## diab_lipids_10          -0.0845562341 -0.015174824 -0.038247213 -0.039466323
## mammogram_10            -0.0086722020 -0.075761844 -0.252186967 -0.266605405
## cs00_seg_inc            -0.1335261996  0.129903297  0.124869466 -0.146697941
## cs00_seg_inc_pov25      -0.1423238073  0.137789729  0.093301631 -0.135737920
## cs00_seg_inc_aff75      -0.1172472439  0.109645355  0.152279349 -0.152756554
## cs_race_theil_2000      -0.0187716853  0.235081082  0.088339010  0.075999576
## gini99                  -0.1536091791  0.451875096  0.200953851  0.208559091
## poor_share              -0.0606871504  0.460239253  0.252434091  0.546190272
## inc_share_1perc         -0.0475465089  0.107960983  0.089184328 -0.040194172
## frac_middleclass         0.2222007066 -0.610306621 -0.242622201 -0.324846689
## scap_ski90pcm            0.4078918674 -0.392857150 -0.210859806 -0.341012495
## rel_tot                  1.0000000000 -0.183567785  0.136445547 -0.221156293
## cs_frac_black           -0.1835677852  1.000000000 -0.110256311  0.267403130
## cs_frac_hisp             0.1364455475 -0.110256311  1.000000000  0.157538376
## unemp_rate              -0.2211562926  0.267403130  0.157538376  1.000000000
## cs_labforce              0.0738993879 -0.278334968 -0.117238150 -0.538895582
## cs_elf_ind_man          -0.0437789796  0.128396588 -0.298207518  0.182485568
## cs_born_foreign         -0.0260191863 -0.051352540  0.660614347  0.052956117
## mig_inflow              -0.2992096491  0.025163787  0.018540568 -0.204919795
## mig_outflow             -0.2429107396  0.116063141  0.089603714 -0.180141961
## pop_density              0.0066804131  0.081106353  0.075521380 -0.010317303
## frac_traveltime_lt15     0.4152105199 -0.277858323  0.167693330 -0.082769827
## hhinc00                  0.0319008919 -0.231116557 -0.108579935 -0.449502685
## median_house_value      -0.2508771864 -0.084624299  0.044705180 -0.182196959
## ccd_exp_tot              0.0090768085 -0.175413783  0.060340388 -0.135214721
## score_r                  0.2170848087 -0.511419257 -0.148645916 -0.292567538
## cs_fam_wkidsinglemom    -0.2401580163  0.803716480  0.009043294  0.390721489
## subcty_exp_pc            0.0812982080 -0.076591124  0.180529958 -0.044063783
## taxrate                  0.0975362196 -0.127895705  0.287385730 -0.167824906
## tax_st_diff_top20        0.0387393759 -0.078949485 -0.010706212 -0.014202824
## deathspc                -0.0365551676  0.282525594 -0.002067729  0.014739487
##                          cs_labforce cs_elf_ind_man cs_born_foreign
## intersects_msa           0.232180037     0.11560522    0.1381784715
## cur_smoke_q1             0.131421941     0.12982235   -0.0171757976
## cur_smoke_q2             0.121716641     0.13046804   -0.0124603997
## cur_smoke_q3             0.070872290     0.12580909   -0.0003196703
## cur_smoke_q4             0.063153298     0.09805783   -0.0109410633
## bmi_obese_q1             0.039697172     0.20388121   -0.0046000491
## bmi_obese_q2             0.051324951     0.19287281   -0.0033034345
## bmi_obese_q3             0.061031687     0.18884617   -0.0216781320
## bmi_obese_q4             0.038663538     0.14680860   -0.0111874702
## exercise_any_q1          0.272141226     0.05500369    0.1356473996
## exercise_any_q2          0.256275691     0.09056310    0.1295170569
## exercise_any_q3          0.250703178     0.09742190    0.1131689082
## exercise_any_q4          0.260842068     0.09696809    0.1292855898
```

```
## brfss_mia              -0.234538464    -0.12184333   -0.1072232251
## puninsured2010         -0.449901540    -0.15857059    0.2774416357
## reimb_penroll_adj10    -0.345372203     0.09776833   -0.0225793101
## mort_30day_hosp_z      -0.198792693     0.11482551   -0.0720371696
## adjmortmeas_amiall30day -0.201237907     0.08089648   -0.0675469992
## adjmortmeas_chfall30day  0.005905577     0.03342141   -0.0430568617
## med_prev_qual_z          0.388507471     0.07401617    0.0344047304
## primcarevis_10         -0.092880152     0.12747697   -0.1719785952
## diab_hemotest_10         0.221667037     0.17863361   -0.0442421799
## diab_eyeexam_10          0.372360764    -0.06831635    0.0301168001
## diab_lipids_10           0.095042298     0.17431791    0.0766547419
## mammogram_10             0.365582667     0.03390542   -0.0137273645
## cs00_seg_inc             0.337104583    -0.15123452    0.3974897207
## cs00_seg_inc_pov25       0.326407311    -0.14213576    0.3390691573
## cs00_seg_inc_aff75       0.336496794    -0.15857893    0.4264072267
## cs_race_theil_2000       0.012486263    -0.07076025    0.2580684489
## gini99                 -0.236239468    -0.12286022    0.2960231218
## poor_share             -0.691802562    -0.10052400    0.0267609656
## inc_share_1perc          0.112076449    -0.07427018    0.2798907526
## frac_middleclass         0.326379970     0.09375352   -0.2675342131
## scap_ski90pcm            0.332591574    -0.07913416   -0.1885520291
## rel_tot                  0.073899388    -0.04377898   -0.0260191863
## cs_frac_black          -0.278334968     0.12839659   -0.0513525402
## cs_frac_hisp           -0.117238150    -0.29820752    0.6606143473
## unemp_rate             -0.538895582     0.18248557    0.0529561166
## cs_labforce              1.000000000     0.08113008    0.1240077465
## cs_elf_ind_man           0.081130077     1.00000000   -0.2443382329
## cs_born_foreign          0.124007746    -0.24433823    1.0000000000
## mig_inflow               0.384551110    -0.07775734    0.2202488899
## mig_outflow              0.408207802    -0.08700974    0.3243495094
## pop_density              0.046675001    -0.06479530    0.3315714110
## frac_traveltime_lt15     0.005329458    -0.24853638   -0.0334308092
## hhinc00                  0.595135410    -0.09280045    0.2482969539
## median_house_value       0.419162730    -0.13613133    0.4387552366
## ccd_exp_tot              0.147686614    -0.19526755    0.1561499374
## score_r                  0.208991059    -0.04617955   -0.2376560858
## cs_fam_wkidsinglemom   -0.352085502     0.05864973    0.0126039266
## subcty_exp_pc            0.197358351    -0.21795101    0.2732350834
## taxrate                  0.153829045    -0.34635102    0.2459051340
## tax_st_diff_top20        0.054308703    -0.07104741    0.1152904420
## deathspc               -0.003612876     0.02274541    0.1754692108
##                          mig_inflow  mig_outflow   pop_density
## intersects_msa           0.452631709  0.440223440  0.0926501872
## cur_smoke_q1             0.225342025  0.195595346  0.0295356049
## cur_smoke_q2             0.228554675  0.179859803  0.0401992956
## cur_smoke_q3             0.176600893  0.144059499  0.0377317694
## cur_smoke_q4             0.164952591  0.130088760  0.0276343168
## bmi_obese_q1             0.169544293  0.133855772  0.0129551022
## bmi_obese_q2             0.177959135  0.137338594 -0.0002068472
## bmi_obese_q3             0.136858480  0.111609914 -0.0037372830
## bmi_obese_q4             0.130027981  0.101085351 -0.0118814735
## exercise_any_q1          0.334210601  0.300677289  0.0619327723
## exercise_any_q2          0.332023565  0.303908423  0.0625316639
## exercise_any_q3          0.319427195  0.282122166  0.0595573463
```

```
## exercise_any_q4          0.326737938  0.286672417  0.0654847025
## brfss_mia                -0.319990830 -0.280927547 -0.0622999946
## puninsured2010           -0.098503059 -0.055554514 -0.0582072386
## reimb_penroll_adj10       0.003512911  0.001013416  0.0160936640
## mort_30day_hosp_z        -0.133248188 -0.122422455 -0.0926498013
## adjmortmeas_amiall30day -0.142037358 -0.125353247 -0.0684475955
## adjmortmeas_chfall30day -0.088842802 -0.092371179 -0.0961238266
## med_prev_qual_z           0.189602583  0.119758304  0.0148901531
## primcarevis_10           -0.034224173 -0.062983026 -0.1131919475
## diab_hemotest_10          0.059695410  0.001239106 -0.0053744821
## diab_eyeexam_10           0.053306161  0.027634786  0.0342012990
## diab_lipids_10            0.199217463  0.119902601  0.0379318265
## mammogram_10              0.152695661  0.079097751  0.0263207356
## cs00_seg_inc              0.384402552  0.452091730  0.2263031757
## cs00_seg_inc_pov25        0.361332747  0.428120122  0.2073271390
## cs00_seg_inc_aff75        0.388634824  0.453580666  0.2273481395
## cs_race_theil_2000        0.094375681  0.165344028  0.2245941279
## gini99                    0.077330297  0.124521937  0.2388933161
## poor_share               -0.349071849 -0.242802629  0.0132631394
## inc_share_1perc           0.114320728  0.148069937  0.2379743130
## frac_middleclass         -0.134202487 -0.189497468 -0.1219529763
## scap_ski90pcm            -0.300350738 -0.286854483 -0.0172066954
## rel_tot                  -0.299209649 -0.242910740  0.0066804131
## cs_frac_black             0.025163787  0.116063141  0.0811063525
## cs_frac_hisp              0.018540568  0.089603714  0.0755213797
## unemp_rate               -0.204919795 -0.180141961 -0.0103173030
## cs_labforce               0.384551110  0.408207802  0.0466750014
## cs_elf_ind_man           -0.077757335 -0.087009738 -0.0647953016
## cs_born_foreign           0.220248890  0.324349509  0.3315714110
## mig_inflow                1.000000000  0.840433024  0.0927415514
## mig_outflow               0.840433024  1.000000000  0.1888314612
## pop_density               0.092741551  0.188831461  1.0000000000
## frac_traveltime_lt15     -0.491195633 -0.421923467 -0.1483695231
## hhinc00                   0.412498140  0.339547198  0.1810464224
## median_house_value        0.434190743  0.438442469  0.4281932633
## ccd_exp_tot              -0.039417807 -0.017785118  0.1049875116
## score_r                  -0.146537646 -0.193660235 -0.1321296359
## cs_fam_wkidsinglemom     -0.044695174  0.065026479  0.1638154079
## subcty_exp_pc            -0.005632313  0.035047881  0.0817929241
## taxrate                  -0.055001250 -0.011748016  0.0215674003
## tax_st_diff_top20        -0.055390482 -0.043939203 -0.0014371559
## deathspc                  0.052352211  0.111084719  0.4174125863
##                         frac_traveltime_lt15       hhinc00 median_house_value
## intersects_msa                  -0.493433447  0.2574416009        0.281379447
## cur_smoke_q1                    -0.214749422  0.0741306347        0.149351428
## cur_smoke_q2                    -0.244463248  0.0918763190        0.158881397
## cur_smoke_q3                    -0.198387924  0.0384608081        0.089493025
## cur_smoke_q4                    -0.181182122  0.0375885600        0.077180115
## bmi_obese_q1                    -0.214271697 -0.0153095129        0.088618795
## bmi_obese_q2                    -0.198747218 -0.0228336766        0.064011768
## bmi_obese_q3                    -0.157411448 -0.0142940407        0.049756874
## bmi_obese_q4                    -0.126294346 -0.0340929940        0.035258512
## exercise_any_q1                 -0.165245916  0.2242189017        0.348497624
## exercise_any_q2                 -0.198362976  0.2044936239        0.328175929
```

```
## exercise_any_q3              -0.199228503  0.2071777939        0.325077972
## exercise_any_q4              -0.209953342  0.2277697402        0.335714908
## brfss_mia                     0.222688379 -0.1819841912       -0.301011917
## puninsured2010                0.056300871 -0.4233601513       -0.241131707
## reimb_penroll_adj10          -0.265178830 -0.2663529324       -0.238039215
## mort_30day_hosp_z             0.092853893 -0.2761392820       -0.217271238
## adjmortmeas_amiall30day       0.087424673 -0.2465907991       -0.223939328
## adjmortmeas_chfall30day       0.190753764 -0.1073923314       -0.088972651
## med_prev_qual_z              -0.058831916  0.4630890270        0.265366704
## primcarevis_10               -0.060976549 -0.1088606725       -0.177416439
## diab_hemotest_10             -0.082109742  0.2616006635        0.100942333
## diab_eyeexam_10               0.186136341  0.3830075350        0.156691819
## diab_lipids_10               -0.295843671  0.2375166539        0.166380805
## mammogram_10                 -0.027252249  0.4668155533        0.294818503
## cs00_seg_inc                 -0.236148771  0.4355155506        0.426761006
## cs00_seg_inc_pov25           -0.215786691  0.3736109834        0.385256953
## cs00_seg_inc_aff75           -0.232502506  0.4720924878        0.441207873
## cs_race_theil_2000           -0.140471968  0.1027073157        0.153490315
## gini99                       -0.175156953 -0.0007508972        0.198526903
## poor_share                    0.061698703 -0.7161368461       -0.380819894
## inc_share_1perc              -0.081784149  0.3206122393        0.415356967
## frac_middleclass              0.334687619  0.0558073603       -0.166657999
## scap_ski90pcm                 0.489457126  0.3311175251        0.008797823
## rel_tot                       0.415210520  0.0319008919       -0.250877186
## cs_frac_black                -0.277858323 -0.2311165569       -0.084624299
## cs_frac_hisp                  0.167693330 -0.1085799354        0.044705180
## unemp_rate                   -0.082769827 -0.4495026849       -0.182196959
## cs_labforce                   0.005329458  0.5951354104        0.419162730
## cs_elf_ind_man               -0.248536380 -0.0928004544       -0.136131329
## cs_born_foreign              -0.033430809  0.2482969539        0.438755237
## mig_inflow                   -0.491195633  0.4124981400        0.434190743
## mig_outflow                  -0.421923467  0.3395471984        0.438442469
## pop_density                  -0.148369523  0.1810464224        0.428193263
## frac_traveltime_lt15          1.000000000 -0.1634064511       -0.259056499
## hhinc00                      -0.163406451  1.0000000000        0.662049953
## median_house_value           -0.259056499  0.6620499528        1.000000000
## ccd_exp_tot                   0.120906284  0.2482094354        0.174112745
## score_r                       0.326764009  0.0007652499       -0.156082814
## cs_fam_wkidsinglemom         -0.146832332 -0.2798330476       -0.074141668
## subcty_exp_pc                 0.264943072  0.2413313105        0.260461559
## taxrate                       0.313515664  0.1716312834        0.129356938
## tax_st_diff_top20             0.059058523  0.0820460575        0.134003939
## deathspc                     -0.180390772  0.1212954036        0.235663948
##                       ccd_exp_tot       score_r cs_fam_wkidsinglemom
## intersects_msa       -0.010579817 -0.1351505704          0.062753741
## cur_smoke_q1         -0.102852758 -0.0706248436          0.115573074
## cur_smoke_q2         -0.082907176 -0.0824370637          0.110207326
## cur_smoke_q3         -0.082393474 -0.0824476086          0.126176976
## cur_smoke_q4         -0.091511257 -0.0697968027          0.070502633
## bmi_obese_q1         -0.131274569 -0.1084552069          0.191762489
## bmi_obese_q2         -0.140022731 -0.0823851842          0.148296845
## bmi_obese_q3         -0.121272646 -0.0590111846          0.112730659
## bmi_obese_q4         -0.115148918 -0.0338788139          0.071447556
## exercise_any_q1      -0.052440103 -0.0607142702          0.083480386
```

25

```
## exercise_any_q2         -0.061409589 -0.0737541592          0.095938908
## exercise_any_q3         -0.055274745 -0.0737442201          0.092184411
## exercise_any_q4         -0.050267103 -0.0768341649          0.099638139
## brfss_mia                0.082339643  0.0946590046         -0.115091618
## puninsured2010          -0.159423970 -0.2307291709          0.199470148
## reimb_penroll_adj10     -0.182632484 -0.1503122982          0.171407957
## mort_30day_hosp_z       -0.127295040 -0.0751753621          0.113858760
## adjmortmeas_amiall30day -0.109129962 -0.0307447156          0.061563075
## adjmortmeas_chfall30day -0.039079969  0.0595916952         -0.079196110
## med_prev_qual_z          0.117845901  0.1704183903         -0.164008443
## primcarevis_10          -0.102755947 -0.0225766254          0.073092667
## diab_hemotest_10         0.057855931  0.1774029631         -0.186427359
## diab_eyeexam_10          0.156136461  0.2362469556         -0.181105404
## diab_lipids_10           0.005154400  0.0377458035         -0.070352091
## mammogram_10             0.151471867  0.1525563057         -0.115130276
## cs00_seg_inc             0.055376696 -0.2059009564          0.249385849
## cs00_seg_inc_pov25       0.052939559 -0.1809964296          0.262718084
## cs00_seg_inc_aff75       0.055235000 -0.2102063397          0.219939875
## cs_race_theil_2000       0.020683675 -0.2878413851          0.377836079
## gini99                  -0.064909097 -0.4421586040          0.574030915
## poor_share              -0.172462323 -0.3525340544          0.581383375
## inc_share_1perc          0.053841388 -0.1803210102          0.170123285
## frac_middleclass         0.068278119  0.5554153128         -0.634387628
## scap_ski90pcm            0.216378601  0.3906793191         -0.365209559
## rel_tot                  0.009076808  0.2170848087         -0.240158016
## cs_frac_black           -0.175413783 -0.5114192567          0.803716480
## cs_frac_hisp             0.060340388 -0.1486459156          0.009043294
## unemp_rate              -0.135214721 -0.2925675380          0.390721489
## cs_labforce              0.147686614  0.2089910589         -0.352085502
## cs_elf_ind_man          -0.195267551 -0.0461795477          0.058649734
## cs_born_foreign          0.156149937 -0.2376560858          0.012603927
## mig_inflow              -0.039417807 -0.1465376457         -0.044695174
## mig_outflow             -0.017785118 -0.1936602351          0.065026479
## pop_density              0.104987512 -0.1321296359          0.163815408
## frac_traveltime_lt15     0.120906284  0.3267640092         -0.146832332
## hhinc00                  0.248209435  0.0007652499         -0.279833048
## median_house_value       0.174112745 -0.1560828135         -0.074141668
## ccd_exp_tot              1.000000000  0.1256517576         -0.119147015
## score_r                  0.125651758  1.0000000000         -0.567812426
## cs_fam_wkidsinglemom    -0.119147015 -0.5678124259          1.000000000
## subcty_exp_pc            0.306078001 -0.0290735403          0.003036160
## taxrate                  0.471192082  0.1519979412         -0.115121884
## tax_st_diff_top20        0.012559877 -0.0574327315         -0.061285871
## deathspc                 0.086697002 -0.1969874736          0.267381528
##                         subcty_exp_pc      taxrate tax_st_diff_top20
## intersects_msa           -0.060726019 -0.095147377      -0.005187465
## cur_smoke_q1             -0.070138557 -0.161371988      -0.051698663
## cur_smoke_q2             -0.043861509 -0.150469778      -0.029537063
## cur_smoke_q3             -0.037991587 -0.107366816      -0.033079094
## cur_smoke_q4             -0.070768586 -0.125859404      -0.031750077
## bmi_obese_q1             -0.065287052 -0.175339820      -0.004540564
## bmi_obese_q2             -0.076833516 -0.175994982       0.002972918
## bmi_obese_q3             -0.063705498 -0.152480688      -0.024956112
## bmi_obese_q4             -0.057609858 -0.136384966       0.004805917
```

```
## exercise_any_q1          0.028602554 -0.097511507       0.030789450
## exercise_any_q2          0.015069532 -0.117209184       0.021033315
## exercise_any_q3          0.015651298 -0.118555933       0.027217036
## exercise_any_q4          0.024200065 -0.104200914       0.025087986
## brfss_mia                0.010208698  0.144250864      -0.020460431
## puninsured2010          -0.060613384  0.056256946      -0.229539579
## reimb_penroll_adj10     -0.216208754 -0.115412915      -0.137556826
## mort_30day_hosp_z       -0.076202926 -0.096191238      -0.036707309
## adjmortmeas_amiall30day -0.086148806 -0.066716196      -0.045117875
## adjmortmeas_chfall30day  0.007297579 -0.031926399       0.009887240
## med_prev_qual_z          0.052046373  0.060840183       0.088846217
## primcarevis_10          -0.175826037 -0.091366236      -0.097031266
## diab_hemotest_10        -0.010876380  0.008928511       0.037649435
## diab_eyeexam_10          0.097899410  0.101894008       0.138399151
## diab_lipids_10          -0.086138784 -0.063381935       0.032153050
## mammogram_10             0.112163964  0.075261910       0.121740299
## cs00_seg_inc             0.167446977  0.043831199       0.057622695
## cs00_seg_inc_pov25       0.156006524  0.039881672       0.044559157
## cs00_seg_inc_aff75       0.173560701  0.053323863       0.063379971
## cs_race_theil_2000       0.122076479  0.011209487       0.054435915
## gini99                   0.047366133 -0.008968813      -0.045718032
## poor_share              -0.136497380 -0.122949406      -0.052101904
## inc_share_1perc          0.149532757  0.064301590       0.013880922
## frac_middleclass         0.046582513  0.062216826       0.120314598
## scap_ski90pcm            0.198925956  0.178687645       0.177558870
## rel_tot                  0.081298208  0.097536220       0.038739376
## cs_frac_black           -0.076591124 -0.127895705      -0.078949485
## cs_frac_hisp             0.180529958  0.287385730      -0.010706212
## unemp_rate              -0.044063783 -0.167824906      -0.014202824
## cs_labforce              0.197358351  0.153829045       0.054308703
## cs_elf_ind_man          -0.217951010 -0.346351021      -0.071047408
## cs_born_foreign          0.273235083  0.245905134       0.115290442
## mig_inflow              -0.005632313 -0.055001250      -0.055390482
## mig_outflow              0.035047881 -0.011748016      -0.043939203
## pop_density              0.081792924  0.021567400      -0.001437156
## frac_traveltime_lt15     0.264943072  0.313515664       0.059058523
## hhinc00                  0.241331310  0.171631283       0.082046057
## median_house_value       0.260461559  0.129356938       0.134003939
## ccd_exp_tot              0.306078001  0.471192082       0.012559877
## score_r                 -0.029073540  0.151997941      -0.057432732
## cs_fam_wkidsinglemom     0.003036160 -0.115121884      -0.061285871
## subcty_exp_pc            1.000000000  0.477105182       0.128411313
## taxrate                  0.477105182  1.000000000      -0.008493485
## tax_st_diff_top20        0.128411313 -0.008493485       1.000000000
## deathspc                 0.037783973  0.019672491       0.002459679
##                             deathspc
## intersects_msa          0.1380284107
## cur_smoke_q1            0.0353149037
## cur_smoke_q2            0.0507259309
## cur_smoke_q3            0.0709302683
## cur_smoke_q4            0.0408965872
## bmi_obese_q1            0.0636890154
## bmi_obese_q2            0.0464543106
## bmi_obese_q3            0.0018303958
```

27

```
## bmi_obese_q4              0.0166350317
## exercise_any_q1           0.0582331912
## exercise_any_q2           0.0780129920
## exercise_any_q3           0.0660440980
## exercise_any_q4           0.0725888164
## brfss_mia                -0.0804931241
## puninsured2010           -0.0441494404
## reimb_penroll_adj10       0.0761997760
## mort_30day_hosp_z        -0.0169124674
## adjmortmeas_amiall30day -0.0192277285
## adjmortmeas_chfall30day -0.0715692421
## med_prev_qual_z          -0.0146153481
## primcarevis_10           -0.0428328977
## diab_hemotest_10         -0.0767657457
## diab_eyeexam_10          -0.0046473518
## diab_lipids_10           -0.0001429391
## mammogram_10              0.0318108456
## cs00_seg_inc              0.2117108671
## cs00_seg_inc_pov25        0.1889817651
## cs00_seg_inc_aff75        0.2109412844
## cs_race_theil_2000        0.2223443325
## gini99                    0.2151956725
## poor_share                0.0807910719
## inc_share_1perc           0.1506526922
## frac_middleclass         -0.2762877717
## scap_ski90pcm            -0.1138163392
## rel_tot                  -0.0365551676
## cs_frac_black             0.2825255942
## cs_frac_hisp             -0.0020677288
## unemp_rate                0.0147394868
## cs_labforce              -0.0036128758
## cs_elf_ind_man            0.0227454143
## cs_born_foreign           0.1754692108
## mig_inflow                0.0523522115
## mig_outflow               0.1110847190
## pop_density               0.4174125863
## frac_traveltime_lt15     -0.1803907721
## hhinc00                   0.1212954036
## median_house_value        0.2356639478
## ccd_exp_tot               0.0866970022
## score_r                  -0.1969874736
## cs_fam_wkidsinglemom      0.2673815284
## subcty_exp_pc             0.0377839731
## taxrate                   0.0196724914
## tax_st_diff_top20         0.0024596793
## deathspc                  1.0000000000
```

```r
# Extract upper triangle of correlation matrix (excluding diagonal)
upper_tri <- cor_matrix[upper.tri(cor_matrix)]
# Compute the average absolute correlation
avg_cor <- mean(abs(upper_tri), na.rm = TRUE)
# Print the result
cat("Average Absolute Correlation:", round(avg_cor, 3), "\n")
```

```
## Average Absolute Correlation: 0.176
```

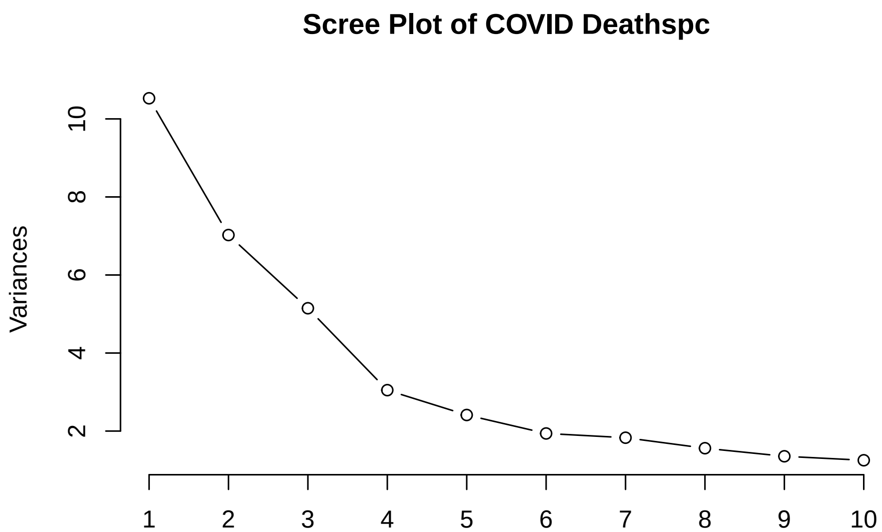
28

```r
cor_df <- as.data.frame(as.table(cor_matrix)) %>%
filter(Var1 != Var2) %>%
arrange(desc(abs(Freq))) # Sort by absolute correlation
head(cor_df, 20) # Show top 10 highest correlations
```

```
##                 Var1              Var2       Freq
## 1  cs00_seg_inc_aff75       cs00_seg_inc  0.9796126
## 2        cs00_seg_inc cs00_seg_inc_aff75  0.9796126
## 3  cs00_seg_inc_pov25       cs00_seg_inc  0.9732649
## 4        cs00_seg_inc cs00_seg_inc_pov25  0.9732649
## 5           brfss_mia    exercise_any_q4 -0.9318503
## 6     exercise_any_q4          brfss_mia -0.9318503
## 7           brfss_mia    exercise_any_q3 -0.9167388
## 8     exercise_any_q3          brfss_mia -0.9167388
## 9  cs00_seg_inc_aff75 cs00_seg_inc_pov25  0.9148627
## 10 cs00_seg_inc_pov25 cs00_seg_inc_aff75  0.9148627
## 11          brfss_mia    exercise_any_q2 -0.9097773
## 12    exercise_any_q2          brfss_mia -0.9097773
## 13    exercise_any_q3    exercise_any_q2  0.8871552
## 14    exercise_any_q2    exercise_any_q3  0.8871552
## 15    exercise_any_q4    exercise_any_q2  0.8814208
## 16    exercise_any_q2    exercise_any_q4  0.8814208
## 17    exercise_any_q2    exercise_any_q1  0.8801893
## 18    exercise_any_q1    exercise_any_q2  0.8801893
## 19    exercise_any_q4    exercise_any_q3  0.8748561
## 20    exercise_any_q3    exercise_any_q4  0.8748561
```

not seeing super strong correlations between covariates

```r
# PCA ANALYSIS
pca_result <- prcomp(df_numeric, scale = TRUE)
plot(pca_result, type = "lines", main = "Scree Plot of COVID Deathspc")
```

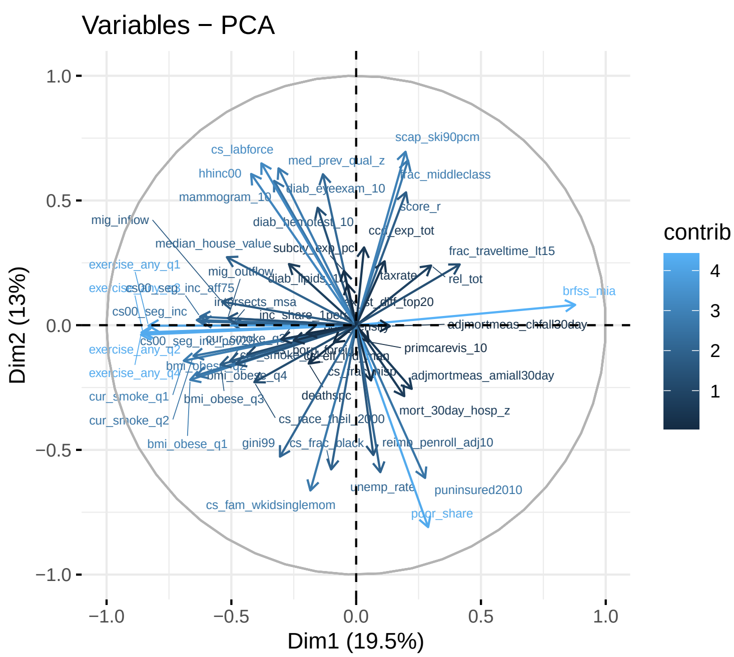



```r
# print PCA significance
print(summary(pca_result))
```

```
## Importance of components:
##                          PC1    PC2     PC3     PC4     PC5    PC6     PC7
## Standard deviation     3.245 2.6502 2.26879 1.74601 1.55302 1.3924 1.35272
## Proportion of Variance 0.195 0.1301 0.09532 0.05645 0.04466 0.0359 0.03389
## Cumulative Proportion  0.195 0.3250 0.42036 0.47682 0.52148 0.5574 0.59127
##                           PC8     PC9    PC10   PC11    PC12    PC13    PC14
## Standard deviation     1.2492 1.16353 1.11884 1.0974 1.03209 0.97409 0.94142
## Proportion of Variance 0.0289 0.02507 0.02318 0.0223 0.01973 0.01757 0.01641
## Cumulative Proportion  0.6202 0.64524 0.66842 0.6907 0.71045 0.72802 0.74443
##                           PC15    PC16    PC17    PC18    PC19    PC20    PC21
## Standard deviation     0.90530 0.89219 0.85496 0.83749 0.81609 0.79522 0.78202
## Proportion of Variance 0.01518 0.01474 0.01354 0.01299 0.01233 0.01171 0.01133
## Cumulative Proportion  0.75961 0.77435 0.78789 0.80088 0.81321 0.82492 0.83625
##                           PC22    PC23    PC24    PC25    PC26    PC27    PC28
## Standard deviation     0.77353 0.75838 0.73258 0.71114 0.70683 0.68869 0.67679
## Proportion of Variance 0.01108 0.01065 0.00994 0.00937 0.00925 0.00878 0.00848
## Cumulative Proportion  0.84733 0.85798 0.86792 0.87728 0.88653 0.89532 0.90380
##                           PC29    PC30    PC31    PC32    PC33    PC34    PC35
## Standard deviation     0.64408 0.63487 0.62556 0.60921 0.58301 0.57708 0.55413
## Proportion of Variance 0.00768 0.00746 0.00725 0.00687 0.00629 0.00617 0.00569
## Cumulative Proportion  0.91148 0.91894 0.92619 0.93306 0.93936 0.94553 0.95121
##                           PC36    PC37   PC38    PC39   PC40    PC41    PC42
## Standard deviation     0.54387 0.52662 0.4929 0.47786 0.4707 0.44603 0.40826
## Proportion of Variance 0.00548 0.00514 0.0045 0.00423 0.0041 0.00368 0.00309
## Cumulative Proportion  0.95669 0.96183 0.9663 0.97055 0.9747 0.97834 0.98143
##                           PC43    PC44    PC45    PC46    PC47    PC48    PC49
## Standard deviation     0.38992 0.35092 0.33924 0.32692 0.32210 0.31347 0.30023
## Proportion of Variance 0.00282 0.00228 0.00213 0.00198 0.00192 0.00182 0.00167
## Cumulative Proportion  0.98424 0.98652 0.98865 0.99063 0.99255 0.99437 0.99604
```

```
##                              PC50    PC51    PC52    PC53    PC54
## Standard deviation     0.27089 0.2546 0.21547 0.16315 0.04926
## Proportion of Variance 0.00136 0.0012 0.00086 0.00049 0.00004
## Cumulative Proportion  0.99740 0.9986 0.99946 0.99996 1.00000
```

```
fviz_pca_var(pca_result, col.var = "contrib", repel = TRUE, labelsize = 2)
```



PRINCIPLE COMPONENT REGRESSION

```
# Convert PCA scores to dataframe
pc_scores <- as.data.frame(pca_result$x)
# Add back the dependent variable (COVID deaths per capita)
df_pcr <- cbind(deathspc = df_numeric$deathspc, pc_scores)
# View first few rows
head(df_pcr)
```

```
##    deathspc       PC1        PC2         PC3       PC4        PC5         PC6
## 1 16.548864 -2.602235  0.3315962  0.05124399 2.4835973  1.1052500 -0.68683357
## 2  8.959118 -2.248282 -0.3626031 -0.62291334 0.9399611  0.3801156  0.32353084
## 3  6.609756 -1.163321 -4.5597171  1.52269774 2.0970423 -3.0310743  0.01330925
## 4  6.038192 -0.564809 -3.4684060  1.61895914 2.0721144  0.2168114 -1.52897629
## 5  1.503713 -1.504646 -0.6160068  2.32643547 1.3318356  1.8573534 -1.22680882
## 6 22.268559  4.832246 -6.4310853 -3.40057097 3.6489565 -2.8730155  0.84014518
##           PC7        PC8         PC9       PC10        PC11      PC12
## 1  0.66058655 -0.9688634  0.51128137 0.08823457  0.01539799 1.3961901
## 2  0.16629797  0.1414011 -0.04700241 0.12644665 -1.82546191 0.1451629
## 3 -0.77495105  0.3065717  0.81920630 0.05653644 -0.59842980 0.7837549
```

```
## 4  0.20458264 -0.2971095 -0.40091332 -0.05972538  0.48352233 -0.2420163
## 5 -0.08805344 -0.6681124 -0.76678153 -0.07451383  0.19729866  0.7645736
## 6  1.19925684  1.7640628  2.61465795 -1.20561289 -0.74020132  0.5220469
##          PC13       PC14        PC15       PC16        PC17        PC18
## 1 -0.03216841  0.2090043  0.24911269 -0.4935048 -0.02804531  0.36542591
## 2  0.47577237  0.3347957  0.92773316 -0.6365187  0.29669573 -0.01909662
## 3  0.64707302 -0.7461183 -0.27763974  0.2552684 -0.65417940 -0.01971288
## 4  0.28605643  0.6788723 -0.01080741 -0.6063605 -0.63555694 -0.59375287
## 5  0.15339464  0.4774014  0.54920518 -0.3000625 -0.37087636  0.40498997
## 6 -0.05324344 -1.6686862 -0.07534531 -0.4730073  0.74705003  0.01217063
##         PC19       PC20         PC21        PC22        PC23        PC24
## 1 -0.3623813 -0.2061607 -0.772083142 -0.36763968  0.02615214  0.12861106
## 2 -0.7190902 -0.2073130 -0.705427460  0.27800574 -1.35030039  0.01713365
## 3 -0.1267119 -0.7365986 -0.005932626 -0.70813103  0.76450973 -0.43467279
## 4 -0.6298228 -0.2356206 -0.569680288 -0.03544688 -0.21222190  0.63469158
## 5 -1.2183557  0.4616638 -1.048799793 -0.54476783 -0.64542348 -1.27665943
## 6  0.2559892  1.0532176  0.123662335 -0.26643911  0.66121265  1.17552507
##          PC25       PC26        PC27        PC28      PC29        PC30
## 1 -0.16924669 -0.4550565 -0.01930892  0.13185609 0.1167402  0.08290766
## 2  0.37943794 -0.0231066 -0.06472087 -0.25004112 0.2146860 -0.29871067
## 3 -0.88336115  0.4539086 -0.08636277 -0.41960318 0.6498133  0.04102403
## 4  0.60879786 -0.3091324 -0.54124451 -0.56144507 1.0315165 -0.01487443
## 5 -0.64495488  0.5314687  0.04910037 -0.02179082 0.1791021 -0.44656415
## 6 -0.09516217  1.3722049 -0.81658174 -0.55957287 0.1041149 -0.10009059
##           PC31        PC32        PC33        PC34        PC35      PC36
## 1  0.199357180 -0.23306332 -0.58537781 -0.82075685  0.01193722 0.3009659
## 2 -0.192057695 -0.52558674  0.26101052 -0.04887761  0.31735598 0.2464218
## 3 -0.167770531 -0.24771709  0.05933639  1.27118200  0.57459090 0.7648313
## 4  0.132976795  0.18560893  0.13967584 -0.57081694  0.22724790 0.2044502
## 5 -0.002150837 -0.34424892  0.15595468 -0.23342645  0.54524566 0.4449939
## 6  0.562270933 -0.02053912  0.16644982 -0.32137025 -0.36481555 1.0882762
##          PC37        PC38       PC39        PC40       PC41        PC42
## 1  0.17220180 -0.33462506  0.1515654 -0.04506444  0.1063434 -0.02841751
## 2 -0.15737224  0.03428701  0.1431777 -0.05104870 -0.2820078 -0.24191964
## 3 -0.09475278  0.02881688  0.5917257  0.66678201  0.6869519  0.11154387
## 4  0.52272340 -0.35463826  0.8615705 -0.28857839  0.3908867 -0.03689141
## 5  0.47524835 -0.39924454  0.3163604 -0.16239063  0.3429887  0.29630143
## 6 -0.12129761  0.44682040 -0.6842020  1.27996225  1.0295946  0.16838345
##          PC43        PC44        PC45        PC46        PC47        PC48
## 1 0.215936700  0.08087658 -0.20774858  0.07228881  0.61994149  0.11560971
## 2 0.009182498 -0.14572710  0.37961262  0.04791434  0.40900382 -0.08684874
## 3 0.295572986 -0.46127000  0.10830136 -0.17737115 -0.02181088  0.29457948
## 4 0.558479780 -0.08854918  0.27866601  0.45794580  0.46835095  0.12861331
## 5 0.059902361 -0.03072468 -0.09234886  0.16196113  0.38235466  0.64720614
## 6 0.008901957  0.09762524  0.39980566  0.07942203 -0.79566491 -0.16166871
##         PC49        PC50        PC51        PC52        PC53          PC54
## 1 -0.1084207  0.04231473  0.20186269 0.131252948  0.01585171 -0.1190882740
## 2  0.2823435 -0.19009279 -0.22976293 0.011352867 -0.11627334  0.0199969806
## 3  0.3642483 -0.32072383  0.21193224 0.008728125  0.07042937  0.0004803217
## 4  0.4035053  0.13944983  0.42534339 0.189764099 -0.21433328  0.0402450456
## 5 -0.3074628  0.03806907  0.01590107 0.126791358  0.10315580 -0.0415648707
## 6  0.2100212  0.18932802 -0.23694536 0.058387720  0.14763890  0.1746249887
```

```r
set.seed(421)
# 80% training, 20% testing
trainIndex <- createDataPartition(df_pcr$deathspc, p = 0.8, list = FALSE)
train_data <- df_pcr[trainIndex, ]
test_data <- df_pcr[-trainIndex, ]
# Store AIC values for different numbers of PCs
aic_values <- numeric(10) # Store AIC for first 10 PCs
for (k in 1:10) {
formula <- as.formula(paste("deathspc ~", paste0("PC", 1:k, collapse = "+")))
lm_model <- lm(formula, data = train_data)
aic_values[k] <- AIC(lm_model)
}
# Find best number of PCs (lowest AIC)
best_k <- which.min(aic_values)
cat("\nBest number of PCs by AIC:", best_k, "\n")
```

```
##
## Best number of PCs by AIC: 10
```

```r
##
## Best number of PCs by AIC: 10
# Fit OLS using the best number of PCs from AIC selection
formula_best <- as.formula(paste("deathspc ~", paste0("PC", 1:best_k, collapse = "+")))
pcr_model <- lm(formula_best, data = train_data)
# Print model summary
summary(pcr_model)
```

```
##
## Call:
## lm(formula = formula_best, data = train_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -316.49  -17.54   -4.88    9.46  408.24
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.7003     0.7662  29.628  < 2e-16 ***
## PC1          -3.1519     0.2349 -13.419  < 2e-16 ***
## PC2          -3.2196     0.2922 -11.018  < 2e-16 ***
## PC3          -7.1262     0.3354 -21.249  < 2e-16 ***
## PC4           2.5178     0.4327   5.819 6.74e-09 ***
## PC5          -4.7202     0.4891  -9.651  < 2e-16 ***
## PC6          -0.2273     0.5537  -0.411    0.681
## PC7           7.0994     0.5636  12.596  < 2e-16 ***
## PC8          15.5298     0.6017  25.811  < 2e-16 ***
## PC9         -14.1587     0.6383 -22.182  < 2e-16 ***
## PC10          8.7186     0.6863  12.704  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.01 on 2323 degrees of freedom
## Multiple R-squared:  0.5157, Adjusted R-squared:  0.5136
## F-statistic: 247.4 on 10 and 2323 DF,  p-value: < 2.2e-16
```

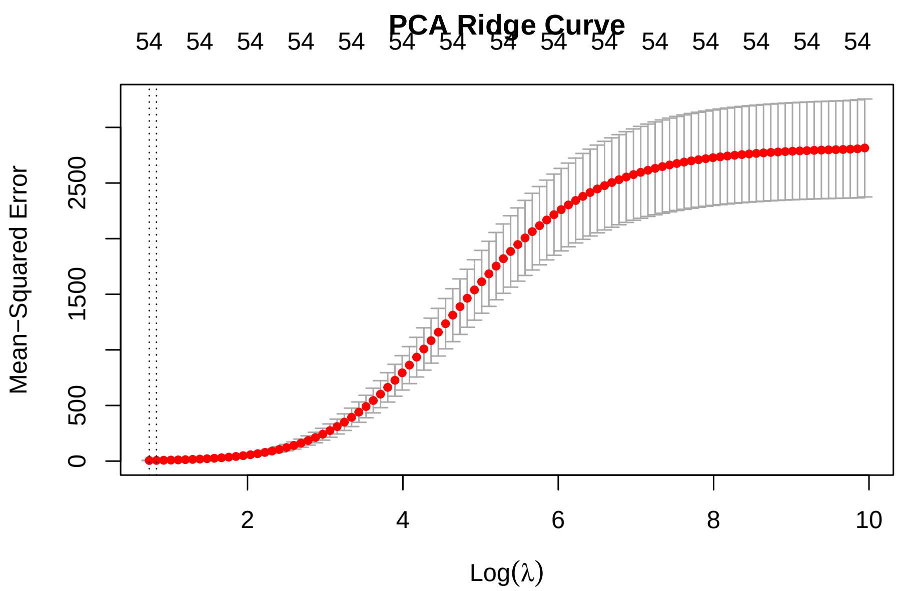
```r
# Make predictions on the test set
pcr_predictions <- predict(pcr_model, newdata = test_data)
# Compute RMSE (Root Mean Squared Error)
rmse <- sqrt(mean((pcr_predictions - test_data$deathspc)^2))
cat("RMSE of PCR Model:", rmse)
```

```
## RMSE of PCR Model: 38.78843
```

```r
## RMSE of PCR Model: 38.78843
# compare this result with Lasso and ridge on original covariates
# Convert PCA-transformed data to matrix (excluding response variable)
x_pca_train <- as.matrix(train_data[, -1]) # PCA scores as predictors
y_train <- train_data$deathspc
# Ridge on PCA scores
cv_ridge_pca <- cv.glmnet(x_pca_train, y_train, alpha = 0)
ridge_lambda_pca <- cv_ridge_pca$lambda.min
cat("pca ridge lambda :", ridge_lambda_pca, "\n")
```

```
## pca ridge lambda : 2.085233
```

```r
plot(cv_ridge_pca, main = "PCA Ridge Curve")
```

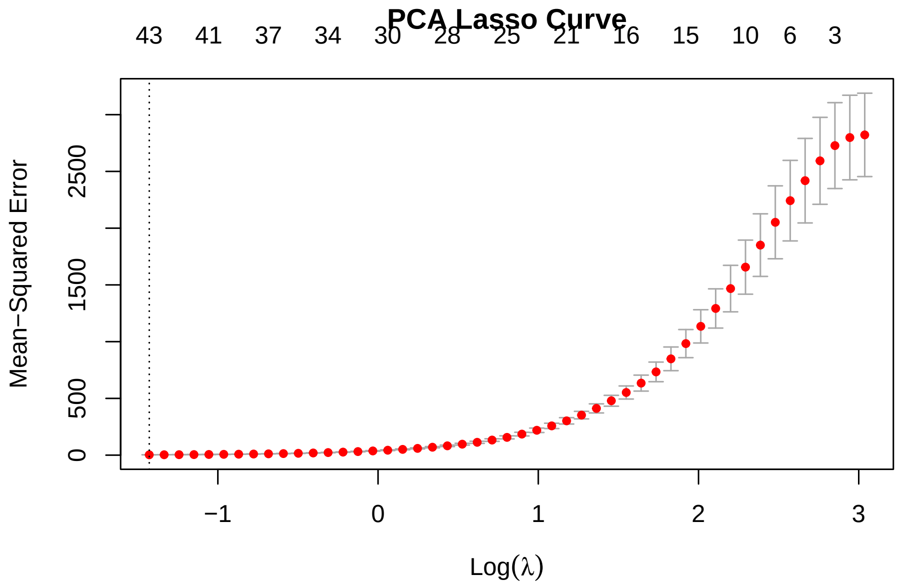



```r
# Lasso on PCA scores
cv_lasso_pca <- cv.glmnet(x_pca_train, y_train, alpha = 1)
lasso_lambda_pca <- cv_lasso_pca$lambda.min
cat("PCA lasso lambda :", lasso_lambda_pca, "\n")
```

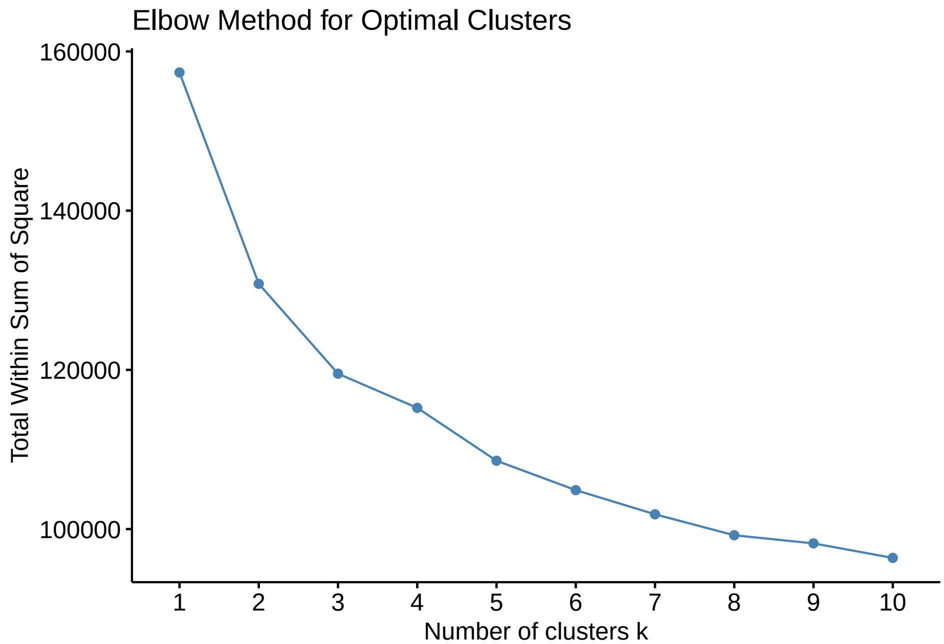
```
## PCA lasso lambda : 0.2397511
```

```r
plot(cv_lasso_pca, main = "PCA Lasso Curve")
```

34




```
# Remove non-numeric columns before clustering
numeric_data <- data_cleaned %>%
  select(-c(county, state))  # Exclude categorical variables
numeric_data <- numeric_data %>%
  mutate(across(everything(), as.numeric))
# Normalize the data (scale to mean 0, variance 1)
numeric_scaled <- scale(numeric_data)

# Determine the optimal number of clusters using the Elbow Method
fviz_nbclust(numeric_scaled, kmeans, method = "wss") +
  labs(title = "Elbow Method for Optimal Clusters")
```




```r
ggsave("Elbow_Method.png", width = 8, height = 6, dpi = 300)

set.seed(421)
kmeans_result <- kmeans(numeric_scaled, centers = 3, nstart = 10)
data_cleaned$cluster <- as.factor(kmeans_result$cluster)
# Calculate mean values for each cluster
cluster_summary <- data_cleaned %>%
  group_by(cluster) %>%
  summarise(
    mean_deathspc = mean(deathspc, na.rm = TRUE),
    mean_poor_share = mean(poor_share, na.rm = TRUE)
  )

# Print the summary to console for reference
print(cluster_summary)
```

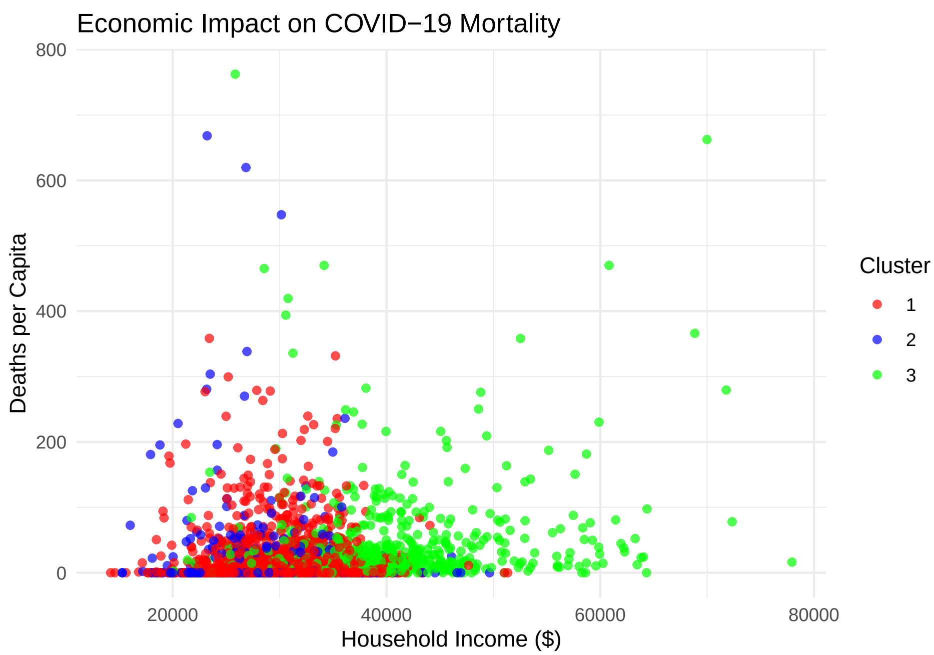
```
## # A tibble: 3 x 3
##   cluster mean_deathspc mean_poor_share
##   <fct>           <dbl>           <dbl>
## 1 1                17.7           0.146
## 2 2                14.8           0.162
## 3 3                40.7           0.111
```

```r
# Optionally, save the summary to a CSV file for record-keeping
write.csv(cluster_summary, "cluster_summary.csv", row.names = FALSE)
```

```r
# Household Income vs. COVID-19 Deaths per Capita
plot1 <- ggplot(data_cleaned, aes(x = hhinc00, y = deathspc, color = cluster)) +
```

36

```r
  geom_point(alpha = 0.7) +
  scale_color_manual(values = c("red", "blue", "green")) +
  labs(title = "Economic Impact on COVID-19 Mortality",
       x = "Household Income ($)", y = "Deaths per Capita", color = "Cluster") +
  theme_minimal()

print(plot1)  # Print plot
```




```r
ggsave("Income_vs_DeathRate.png", plot = plot1, width = 8, height = 6, dpi = 300)
```

Interpretation of Economic Impact on COVID-19 Mortality Clustering: This scatter plot represents the relationship between household income and COVID-19 deaths per capita, with data points grouped into three clusters using K-Means clustering. The X-axis represents household income, while the Y-axis represents COVID-19 deaths per capita. The different colors denote clusters: Cluster 1 (Red): Counties with high mortality rates, widely distributed across different income levels. Cluster 2 (Blue): Counties with moderate mortality rates, concentrated mostly in lower-to-middle-income ranges. Cluster 3 (Green): Counties with lower mortality rates, primarily located in middle-income groups. Key Insights: 1. Higher-income does not always mean lower mortality: Unlike the previous plot, Cluster 1 (Red, high mortality) includes many counties in the higher-income bracket ($40,000 - $70,000). This suggests that factors beyond income, such as healthcare access, comorbidities, and regional pandemic policies, are influencing mortality rates. 2. Lower-income counties still show significant mortality risks: Cluster 2 (Blue) contains many low-to-middle-income counties ($15,000 - $40,000) with moderate death rates, reinforcing that economic disadvantage contributes to COVID-19 severity but is not the sole determinant. 3. Counties in Cluster 3 (Green, low mortality) are mostly concentrated at the lower range of deaths per capita: These counties are spread across different income levels but tend to have better overall health outcomes, possibly due to lower population density, stronger healthcare infrastructure, or better public health policies. Policy Recommendations Based on K-Means Clustering: 1.Target high-mortality counties (Cluster 1) for urgent interventions, regardless of income level.

These counties need enhanced healthcare infrastructure, expanded vaccine access, and better preparedness for future pandemics. 2. Investigate additional risk factors influencing high-income counties with high mortality. Since some wealthier counties still suffer high death rates, factors such as prevalence of pre-existing conditions (obesity, diabetes), urbanization, and healthcare system efficiency should be further analyzed. 3. Provide financial and healthcare support for low-income counties in Cluster 2 (Blue). These counties are at a moderate risk and could benefit from economic assistance, public health awareness programs, and increased medical resource allocation.

```
unique(data_cleaned$bmi_obese_q1)
```

```
##   [1] 0.37500000 0.29805014 0.29411766 0.46666667 0.34782609 0.00000000
##   [7] 0.58333331 0.25000000 0.44444445 0.36363637 0.36000001 0.51948053
##  [13] 0.43333334 0.60000002 0.42857143 0.36956522 0.40983605 0.30000001
##  [19] 0.39024389 0.18750000 0.23076923 0.35714287 0.52482271 0.30303031
##  [25] 0.31343284 0.44117647 0.28467155 0.28571430 0.48387095 0.54054052
##  [31] 0.36538461 0.38461539 0.31912568 0.26400000 0.25270757 0.25806451
##  [37] 0.43243244 0.30227745 0.46551725 0.29787233 0.23255815 0.33251634
##  [43] 0.29729730 0.36115843 0.32571429 0.44680852 0.32499999 0.36458334
##  [49] 0.27500001 0.39004150 0.35185185 0.32928944 0.38693467 0.40000001
##  [55] 0.30769232 0.33437499 0.25248072 0.19184290 0.32579187 0.35064936
##  [61] 0.30927834 0.21069182 0.24381188 0.26333907 0.23903401 0.33531511
##  [67] 0.27261904 0.18695652 0.31738281 0.40243903 0.41860464 0.25769231
##  [73] 0.26418152 0.29608938 0.27272728 0.47058824 0.27380952 0.39534885
##  [79] 0.32307693 0.29675812 0.25652173 0.40721649 0.28787878 0.26315790
##  [85] 0.26744187 0.31538463 0.46153846 0.30252102 0.23750000 0.23529412
##  [91] 0.33333334 0.36246786 0.22916667 0.41176471 0.56250000 0.31372550
##  [97] 0.29496402 0.18181819 0.38812786 0.35227272 0.38834950 0.53846157
## [103] 0.35555556 0.29166666 0.30232558 0.29499322 0.40425533 0.26388890
## [109] 0.31453362 0.31818181 0.34693879 0.34082398 0.37288135 0.21727395
## [115] 0.27154046 0.47727272 0.21682848 0.21428572 0.20000000 0.27480915
## [121] 0.28903654 0.50000000 0.33928570 0.20588236 0.09090909 1.00000000
## [127] 0.24053296 0.26470590 0.20930232 0.10714286 0.31578946 0.26724139
## [133] 0.26666668 0.20759717 0.13513513 0.25531915 0.24887556 0.16666667
## [139] 0.29533678 0.22707424 0.14322251 0.27845037 0.24444444 0.19902913
## [145] 0.24878049 0.24319419 0.28947368 0.25600001 0.23381294 0.33892617
## [151] 0.37185928 0.55555558 0.23312883 0.25331858 0.20833333 0.18400000
## [157] 0.14657980 0.21126761 0.16891892 0.21014114 0.20769231 0.21763754
## [163] 0.16814159 0.14062500 0.16279070 0.21675977 0.21323529 0.19546743
## [169] 0.17142858 0.18666667 0.21746881 0.22857143 0.21794872 0.18309858
## [175] 0.26428571 0.36842105 0.15384616 0.09523810 0.21934369 0.22222222
## [181] 0.07317073 0.19078948 0.27192983 0.25329694 0.26648772 0.20394737
## [187] 0.24494383 0.27033871 0.25743854 0.21991701 0.29883569 0.29575598
## [193] 0.26655897 0.30256251 0.22961374 0.35153583 0.26956522 0.42592594
## [199] 0.21669979 0.22278057 0.32989690 0.25939849 0.32121211 0.26818183
## [205] 0.22190201 0.32870370 0.34065935 0.36312848 0.27871940 0.30315790
## [211] 0.22164948 0.49315068 0.35238096 0.29019609 0.40756303 0.38965517
## [217] 0.38793105 0.27302632 0.22674419 0.29763129 0.32722512 0.29710144
## [223] 0.31983805 0.38493723 0.29846939 0.23366337 0.23923445 0.31034482
## [229] 0.33561644 0.24671917 0.27841845 0.22596154 0.24158500 0.19014084
## [235] 0.33587787 0.27058825 0.34298441 0.24557753 0.24509804 0.22247446
## [241] 0.28070176 0.22233105 0.29520866 0.24117647 0.30630630 0.19724771
## [247] 0.22869022 0.23376623 0.30714285 0.32142857 0.34196892 0.22792023
## [253] 0.32188842 0.32745591 0.35294119 0.31914893 0.25999999 0.40909091
## [259] 0.29184550 0.26086956 0.27777779 0.38571429 0.37837839 0.29674795
## [265] 0.23021583 0.25110132 0.31517509 0.26649076 0.37903225 0.28799999
```

```
## [271] 0.21739130 0.10000000 0.54687500 0.24827586 0.37593985 0.42105263
## [277] 0.36666667 0.38805971 0.27941176 0.28867403 0.54545456 0.36879432
## [283] 0.13333334 0.40740740 0.45454547 0.24535316 0.25555557 0.07142857
## [289] 0.62500000 0.37634408 0.21348314 0.35507247 0.21052632 0.36305732
## [295] 0.10344828 0.44999999 0.14285715 0.54166669 0.32964602 0.58823532
## [301] 0.51515150 0.11111111 0.41666666 0.26923078 0.34999999 0.33268860
## [307] 0.57142860 0.43750000 0.19047619 0.40952381 0.49056605 0.37931034
## [313] 0.30526316 0.08333334 0.25781250 0.22267206 0.22369389 0.39285713
## [319] 0.26515150 0.12068965 0.21531101 0.24010110 0.21153846 0.28300804
## [325] 0.23783784 0.26126125 0.23566879 0.26213592 0.22695035 0.24275362
## [331] 0.26296297 0.27065027 0.28382838 0.25079367 0.24617524 0.18927445
## [337] 0.20238096 0.21573605 0.25905293 0.26226225 0.32075471 0.25227964
## [343] 0.24891210 0.22800000 0.28523862 0.12500000 0.25700936 0.24696356
## [349] 0.26553673 0.26436782 0.29253730 0.71428573 0.28708133 0.33812949
## [355] 0.36708862 0.24712643 0.23404256 0.77777779 0.17647059 0.32862189
## [361] 0.27247956 0.26822916 0.32558140 0.31707317 0.24867725 0.34630349
## [367] 0.07692308 0.32154340 0.34117648 0.36153847 0.21176471 0.34615386
## [373] 0.36029410 0.30065361 0.23584905 0.30434781 0.36040610 0.35135135
## [379] 0.29836065 0.32384342 0.16000000 0.16725978 0.31250000 0.28888890
## [385] 0.52380955 0.27074236 0.27679783 0.61538464 0.27516779 0.29230770
## [391] 0.31018519 0.27408639 0.32530120 0.28135592 0.39130434 0.26249999
## [397] 0.26829270 0.29569891 0.25435540 0.30263159 0.20232558 0.28591749
## [403] 0.53333336 0.30348259 0.66666669 0.28301886 0.27105451 0.30214426
## [409] 0.26355422 0.32051283 0.16169155 0.28205130 0.36477986 0.31927711
## [415] 0.29263565 0.11764706 0.39682540 0.43636364 0.33505154 0.30379745
## [421] 0.27436823 0.33593750 0.33760685 0.27000001 0.36486486 0.18867925
## [427] 0.28275862 0.31081080 0.26760563 0.38383839 0.23853211 0.32777777
## [433] 0.35087720 0.22480001 0.36134455 0.23849373 0.28451884 0.26530612
## [439] 0.31999999 0.36263737 0.28169015 0.24830700 0.27661598 0.33734939
## [445] 0.31202045 0.32710281 0.32167152 0.32692307 0.27530363 0.31168830
## [451] 0.29338843 0.29020333 0.34736842 0.39215687 0.20242915 0.28840971
## [457] 0.33630952 0.27645051 0.26252159 0.24666667 0.28328612 0.38021979
## [463] 0.30350193 0.33175355 0.30866808 0.35443038 0.27956989 0.35962147
## [469] 0.28792384 0.33474576 0.24040404 0.33085501 0.33663365 0.33707866
## [475] 0.35922331 0.32620320 0.30640668 0.29464287 0.26732674 0.26056337
## [481] 0.31428573 0.25742576 0.27200001 0.41999999 0.23636363 0.31901839
## [487] 0.34412956 0.28915662 0.35099337 0.34598213 0.33592737 0.28747794
## [493] 0.47619048 0.32397407 0.36170211 0.25149700 0.15000001 0.23787528
## [499] 0.33910036 0.33187774 0.36111110 0.20710059 0.41463414 0.34722221
## [505] 0.27402136 0.31762296 0.32236096 0.28329486 0.35616440 0.18367347
## [511] 0.32921812 0.45945945 0.30939227 0.43055555 0.26324505 0.36423841
## [517] 0.37113401 0.31314433 0.31811696 0.33553720 0.52499998 0.41379312
## [523] 0.39823008 0.48809522 0.45238096 0.38043478 0.32022473 0.44366196
## [529] 0.39207047 0.26611227 0.33937824 0.34733894 0.40816328 0.34883720
## [535] 0.34375000 0.34146342 0.27350429 0.28148147 0.25794154 0.22488038
## [541] 0.28702012 0.24833703 0.27931961 0.26047358 0.22292994 0.29822162
## [547] 0.27209100 0.33606556 0.30177516 0.28546712 0.29868227 0.28925619
## [553] 0.24638912 0.27298051 0.29380530 0.24700071 0.27053139 0.34188035
## [559] 0.26609442 0.30077121 0.33557048 0.37755102 0.27101201 0.29859155
## [565] 0.28526646 0.24431819 0.38938054 0.23850575 0.45833334 0.34285715
## [571] 0.32804233 0.32038835 0.33224967 0.19649805 0.26118067 0.30228832
## [577] 0.27457955 0.25650558 0.30447942 0.22273782 0.25161290 0.20625000
## [583] 0.25649351 0.27625501 0.25906894 0.37037036 0.34482759 0.30821916
## [589] 0.39772728 0.30612245 0.38235295 0.32258064 0.36559141 0.27586207
```

```
## [595] 0.32695138 0.38095239 0.28846154 0.23999999 0.31473213 0.51428574
## [601] 0.43589744 0.31840795 0.29333332 0.26393628 0.41818181 0.28395063
## [607] 0.26490065 0.36734694 0.34666666 0.31782946 0.38311687 0.27173913
## [613] 0.47368422 0.32936507 0.39416060 0.30733946 0.37142858 0.34328359
## [619] 0.25974026 0.34264559 0.69565219 0.44000000 0.21621622 0.30645162
## [625] 0.23387097 0.26682135 0.75000000 0.25714287 0.28125000 0.23355098
## [631] 0.25196850 0.63636363 0.25688073 0.22153845 0.27397260 0.25352111
## [637] 0.22254336 0.24358974 0.40856031 0.27624309 0.46226415 0.38864627
## [643] 0.34210527 0.35897437 0.40322581 0.46564886 0.47692308 0.34015346
## [649] 0.31328321 0.27027026 0.46428570 0.33027524 0.27329192 0.32452831
## [655] 0.32561612 0.50442475 0.27631578 0.30833334 0.37789202 0.34228188
## [661] 0.27218935 0.34422657 0.48979592 0.30339807 0.44140625 0.38636363
## [667] 0.33467743 0.33673468 0.32786885 0.30737704 0.38827839 0.32584271
## [673] 0.28735632 0.27684963 0.39795917 0.54966885 0.38775510 0.30952382
## [679] 0.26027396 0.31944445 0.38644066 0.42129630 0.37777779 0.24242425
## [685] 0.22727273 0.22955145 0.29878870 0.26016259 0.27963525 0.27840909
## [691] 0.17391305 0.23333333 0.32911393 0.27815467 0.28378379 0.27810651
## [697] 0.24633431 0.34112149 0.06250000 0.14814815 0.31527093 0.38888890
## [703] 0.16463415 0.28853756 0.28279069 0.27927929 0.46575344 0.35403726
## [709] 0.21818182 0.33090910 0.27570093 0.28944382 0.28000000 0.21276596
## [715] 0.31007752 0.21327014 0.14795008 0.30930930 0.27607360 0.16216215
## [721] 0.27352473 0.25118482 0.18918920 0.18732525 0.16444445 0.21158691
## [727] 0.34078214 0.25454545 0.21575984 0.24204947 0.28929386 0.29437229
## [733] 0.29133859 0.13636364 0.27368420 0.33262712 0.20270270 0.29347825
## [739] 0.26974198 0.52941179 0.30830041 0.31850788 0.17857143 0.30882353
## [745] 0.19512194 0.25106642 0.29840848 0.25503355 0.40606061 0.30287206
## [751] 0.28048781 0.27593818 0.31362468 0.28682169 0.19696970 0.34710744
## [757] 0.33076924 0.27419356 0.23083586 0.17445482 0.27160493 0.22911051
## [763] 0.28717950 0.21316226 0.20766129 0.26043406 0.22137405 0.27198917
## [769] 0.29898402 0.26112413 0.25063163 0.24516129 0.23769872 0.27906978
## [775] 0.27378640 0.19748157 0.29436326 0.26599327 0.27879798 0.28449437
## [781] 0.24017467 0.22680412 0.13725491 0.26755852 0.33456790 0.24601063
## [787] 0.20647773 0.27799737 0.26952142 0.22335026 0.15818182 0.26627219
## [793] 0.29545453 0.17298578 0.28455284 0.26891893 0.27469134 0.25862068
## [799] 0.28426397 0.28476822 0.28681320 0.26994255 0.29824561 0.22838138
## [805] 0.22482014 0.28899083 0.30039525 0.25961539 0.24615385 0.29885057
## [811] 0.21061499 0.33009708 0.24731183 0.44537815 0.30188680 0.24713585
## [817] 0.29508197 0.24641834 0.31504703 0.40384614 0.26190478 0.32352942
## [823] 0.45625001 0.51260507 0.27727273 0.24324325 0.29617834 0.31309298
## [829] 0.29181495 0.29104477 0.27426809 0.26737967 0.23214285 0.29319373
## [835] 0.38267148 0.33833718 0.27118644 0.28742516 0.29818782 0.46850392
## [841] 0.33552632 0.28605482 0.42424244 0.26659411 0.40514469 0.36979166
## [847] 0.31443298 0.27246377 0.39102563 0.39834026 0.30831644 0.27230048
## [853] 0.34717607 0.23888889 0.46000001 0.28063241 0.40277779 0.28292683
## [859] 0.35640138 0.41481480 0.31182796 0.38053098 0.37751004 0.28082192
## [865] 0.34751773 0.37824675 0.30484694 0.42307693 0.24852072 0.30705395
## [871] 0.30693069 0.25316456 0.28609625 0.40975609 0.28236607 0.44628099
## [877] 0.23041475 0.44871795 0.34817815 0.36283186 0.30136988 0.22672509
## [883] 0.23342736 0.22965440 0.28400955 0.28037384 0.26424870 0.30851063
## [889] 0.23117709 0.30666667 0.45652175 0.20454545 0.34011090 0.19230770
## [895] 0.44186047 0.30489191 0.48484850 0.32323232 0.39655173 0.26415095
## [901] 0.44736841 0.32478634 0.30756578 0.32638437 0.30581948 0.31631097
## [907] 0.32203391 0.56521738 0.39622641 0.64999998 0.31990924 0.32231405
## [913] 0.25925925 0.32919255 0.31506848 0.29131654 0.31481481 0.32941177
```

```
## [919] 0.24111675 0.29542303 0.31623933 0.26804122 0.23129252 0.30324075
## [925] 0.29052633 0.29770991 0.31468531 0.25704226 0.21212122 0.31410256
## [931] 0.30660376 0.36702126 0.34749034 0.30701753 0.34154931 0.30191973
## [937] 0.27155444 0.32183909 0.29493088 0.30508474 0.25519288 0.33874241
## [943] 0.29249012 0.29217392 0.32957748 0.33640552 0.15942030 0.27175781
## [949] 0.27055702 0.18032786 0.19642857 0.23208191 0.23721436 0.31666666
## [955] 0.33070865 0.20995671 0.28780487 0.24493927 0.24489796 0.26039782
## [961] 0.28033474 0.29106030 0.30046511 0.24667189 0.33177570 0.22000000
## [967] 0.28658536 0.19565217 0.32653061 0.23580366 0.28294572 0.25537634
## [973] 0.32600001 0.28685260 0.28438228 0.30386740 0.30787590 0.31547618
## [979] 0.28923076 0.39560440 0.32885906 0.19718310 0.45762712 0.28630707
## [985] 0.27001861 0.35826477 0.28723404 0.26628897 0.29173419 0.40366971
## [991] 0.24857143 0.31273270 0.30370370 0.35416666 0.29629630 0.25825244
## [997] 0.31228071 0.38281250 0.35986775 0.31683168 0.33636364 0.04347826
## [1003] 0.34682405 0.27802691 0.41935483 0.28276879 0.28409091 0.21483377
## [1009] 0.22972973 0.22349571 0.24674444 0.23240939 0.33043477 0.33705080
## [1015] 0.30328867 0.40196079 0.27620968 0.28737691 0.29473683 0.36015326
## [1021] 0.37119114 0.32450330 0.34494773 0.32291666 0.43103448 0.34508076
## [1027] 0.26770428 0.30372491 0.51200002 0.26561126 0.35436893 0.34554973
## [1033] 0.33759591 0.45161289 0.31200001 0.34959349 0.32070708 0.44101125
## [1039] 0.22823529 0.29559749 0.30270794 0.34280640 0.37209302 0.42372882
## [1045] 0.29465929 0.27207637 0.24561404 0.35675675 0.25252524 0.52857143
## [1051] 0.18803419 0.24778761 0.29487181 0.39199999 0.26881722 0.27049181
## [1057] 0.28648648 0.25217390 0.16049382 0.22807017 0.25835189 0.24350484
## [1063] 0.22741273 0.30120483 0.23711340 0.26178011 0.27131784 0.23460411
## [1069] 0.17073171 0.29600000 0.46875000 0.28598484 0.51724136 0.29032257
## [1075] 0.26015228 0.48148149 0.39220780 0.33649290 0.25641027 0.23188406
## [1081] 0.30143541 0.26863354 0.41538462 0.29084969 0.35106382 0.43181819
## [1087] 0.36783215 0.12000000 0.33699635 0.25694445 0.17241380 0.27140883
## [1093] 0.26395938 0.31946853 0.30172414 0.25477707 0.27845627 0.22413793
## [1099] 0.37449798 0.33703703 0.31106472 0.29816514 0.40625000 0.23756906
## [1105] 0.37391305 0.32432431 0.33653846 0.32042253 0.41025642 0.26363635
## [1111] 0.24285714 0.25164691 0.32158589 0.35483870 0.28985506 0.24468085
## [1117] 0.19526628 0.25060827 0.20436507 0.29325512 0.20481928 0.23196559
## [1123] 0.22317597 0.13432837 0.24025974 0.28429425 0.20400728 0.19030520
## [1129] 0.21666667 0.24904214 0.26682988 0.25548902 0.25189188 0.22041355
## [1135] 0.25301206 0.27677724 0.26484752 0.24871795 0.24713244 0.25186032
## [1141] 0.26390871 0.21044303 0.24075367 0.31698114 0.19166666 0.48275861
## [1147] 0.18248175 0.44897959 0.23157895 0.27710843 0.29007635 0.30107528
## [1153] 0.38345864 0.20338982 0.31092438 0.35256410 0.31132075 0.36697248
## [1159] 0.30208334 0.37804878 0.31052631 0.35888502 0.32033426 0.28337875
## [1165] 0.25691700 0.35619470 0.29273504 0.30308220 0.31971678 0.33168316
## [1171] 0.37972769 0.30717862 0.29149798 0.29722923 0.37362638 0.35384616
## [1177] 0.32753164 0.24938272 0.20440252 0.23828964 0.28003314 0.25659472
## [1183] 0.32929781 0.33511585 0.29425839 0.32087913 0.34081632 0.28656715
## [1189] 0.31784168 0.16141732 0.28909090 0.25146198 0.27144161 0.27976686
## [1195] 0.30476192 0.27974087 0.33163264 0.23566215 0.30227274 0.31389579
## [1201] 0.31900454 0.33980581 0.21917808 0.27659574 0.32265446 0.28716215
## [1207] 0.26543209 0.26398602 0.29299363 0.27083334 0.28308323 0.32970026
## [1213] 0.25732216 0.25773194 0.32407406 0.23801653 0.31884059 0.37823835
## [1219] 0.24063116 0.35164836 0.28807947 0.28592592 0.57575756 0.31395349
## [1225] 0.30726257 0.27164179 0.48780489 0.34090909 0.29268292 0.23428571
## [1231] 0.43396226 0.22480620 0.27358490 0.32061067 0.32335329 0.28512397
## [1237] 0.32746175 0.25581396 0.23684211 0.23880596 0.29773462 0.30545455
```

```
## [1243] 0.34000000 0.40540540 0.29370630 0.42500001 0.18957347 0.18680090
## [1249] 0.23175965 0.29677418 0.27114967 0.27046263 0.18518518 0.24156305
## [1255] 0.27062705 0.20967741 0.22724766 0.23098591 0.26143423 0.21219823
## [1261] 0.21167883 0.24720894 0.12745099 0.28318584
```

```r
sum(data_cleaned$bmi_obese_q1 == 0)
```

```
## [1] 622
```

```r
data_cleaned %>%
  filter(bmi_obese_q1 == 0) %>%
  select(county, state, bmi_obese_q1, deathspc) %>%
  print()
```

```
##                 county          state bmi_obese_q1    deathspc
## 1              Bullock        Alabama            0  22.2685590
## 2                Coosa        Alabama            0  23.8204040
## 3               Greene        Alabama            0  62.8377530
## 4              Lowndes        Alabama            0 113.0598100
## 5                Perry        Alabama            0   0.0000000
## 6               Sumter        Alabama            0  62.1791380
## 7               Wilcox        Alabama            0  79.8405000
## 8             Greenlee        Arizona            0   0.0000000
## 9              Bradley       Arkansas            0  20.3410680
## 10             Calhoun       Arkansas            0   0.0000000
## 11              Chicot       Arkansas            0   0.0000000
## 12           Cleveland       Arkansas            0   0.0000000
## 13              Dallas       Arkansas            0   0.0000000
## 14              Fulton       Arkansas            0   0.0000000
## 15           Lafayette       Arkansas            0  25.6671680
## 16                 Lee       Arkansas            0  24.3453390
## 17              Monroe       Arkansas            0   0.0000000
## 18              Newton       Arkansas            0   0.0000000
## 19               Perry       Arkansas            0   0.0000000
## 20                Pike       Arkansas            0   0.0000000
## 21             Prairie       Arkansas            0   0.0000000
## 22               Scott       Arkansas            0   0.0000000
## 23              Searcy       Arkansas            0   0.0000000
## 24               Stone       Arkansas            0   0.0000000
## 25            Woodruff       Arkansas            0   0.0000000
## 26            Mariposa     California            0   5.3152571
## 27               Modoc     California            0   0.0000000
## 28                Mono     California            0  25.1611120
## 29              Sierra     California            0   0.0000000
## 30             Trinity     California            0   0.0000000
## 31                Bent       Colorado            0   0.0000000
## 32         Clear Creek       Colorado            0  24.7184510
## 33             Conejos       Colorado            0   0.0000000
## 34            Costilla       Colorado            0   0.0000000
## 35             Crowley       Colorado            0  72.5540010
## 36              Custer       Colorado            0   0.0000000
## 37            Huerfano       Colorado            0   0.0000000
## 38          Kit Carson       Colorado            0  70.3172450
## 39             Lincoln       Colorado            0   0.0000000
## 40          Rio Grande       Colorado            0   0.9277772
```

```
## 41          Saguache     Colorado              0   4.5036120
## 42            Summit     Colorado              0  17.8310780
## 43        Washington     Colorado              0   0.0000000
## 44            Glades      Florida              0  27.8938770
## 45         Lafayette      Florida              0   0.0000000
## 46          Atkinson      Georgia              0  35.6125370
## 47             Bacon      Georgia              0  36.1217610
## 48             Baker      Georgia              0 184.5716200
## 49             Banks      Georgia              0   0.9774025
## 50             Butts      Georgia              0 149.1792600
## 51           Calhoun      Georgia              0 195.3616800
## 52           Candler      Georgia              0   0.0000000
## 53          Charlton      Georgia              0   9.7096357
## 54     Chattahoochee      Georgia              0   0.0000000
## 55            Clinch      Georgia              0  18.6485480
## 56          Crawford      Georgia              0   0.0000000
## 57             Dooly      Georgia              0 195.9870100
## 58             Early      Georgia              0 547.4336500
## 59             Evans      Georgia              0   0.0000000
## 60          Glascock      Georgia              0   0.0000000
## 61           Hancock      Georgia              0 180.7278400
## 62             Heard      Georgia              0  38.3351860
## 63             Irwin      Georgia              0  23.0981560
## 64        Jeff Davis      Georgia              0  16.5169660
## 65           Jenkins      Georgia              0  34.8083920
## 66           Johnson      Georgia              0  52.2824170
## 67            Lanier      Georgia              0  45.0358730
## 68           Lincoln      Georgia              0   2.7520158
## 69              Long      Georgia              0   4.5319152
## 70          McIntosh      Georgia              0   0.8544489
## 71            Marion      Georgia              0  41.1401290
## 72            Miller      Georgia              0   0.0000000
## 73        Montgomery      Georgia              0   0.0000000
## 74          Randolph      Georgia              0 668.2133800
## 75            Schley      Georgia              0  68.4546360
## 76          Seminole      Georgia              0  58.5976450
## 77           Stewart      Georgia              0   3.9177661
## 78            Talbot      Georgia              0  45.8195880
## 79            Taylor      Georgia              0  70.9437410
## 80           Terrell      Georgia              0 619.6058300
## 81             Towns      Georgia              0  18.4229930
## 82          Treutlen      Georgia              0   0.0000000
## 83            Turner      Georgia              0 270.0012500
## 84            Twiggs      Georgia              0   0.5869337
## 85            Warren      Georgia              0   0.0000000
## 86           Wheeler      Georgia              0   0.0000000
## 87            Wilcox      Georgia              0 280.5881700
## 88            Wilkes      Georgia              0  10.1403100
## 89         Wilkinson      Georgia              0  73.0823060
## 90             Adams        Idaho              0   0.0000000
## 91         Bear Lake        Idaho              0   0.0000000
## 92            Custer        Idaho              0   0.0000000
## 93             Lewis        Idaho              0   0.0000000
## 94           Lincoln        Idaho              0   0.0000000
```

```
## 95          Oneida      Idaho    0   0.0000000
## 96          Owyhee      Idaho    0   0.0000000
## 97           Power      Idaho    0   0.0000000
## 98           Teton      Idaho    0   0.0000000
## 99          Valley      Idaho    0   0.0000000
## 100      Alexander   Illinois    0   0.0000000
## 101           Bond   Illinois    0  14.4923430
## 102          Brown   Illinois    0   0.0000000
## 103        Calhoun   Illinois    0   0.0000000
## 104     Cumberland   Illinois    0   0.0000000
## 105        Douglas   Illinois    0   0.0000000
## 106        Edwards   Illinois    0   0.0000000
## 107       Gallatin   Illinois    0   0.0000000
## 108       Hamilton   Illinois    0   0.0000000
## 109         Hardin   Illinois    0   0.0000000
## 110      Henderson   Illinois    0   0.0000000
## 111         Jasper   Illinois    0 132.9185500
## 112       Marshall   Illinois    0   0.0000000
## 113         Menard   Illinois    0   0.0000000
## 114       Moultrie   Illinois    0   0.0000000
## 115          Piatt   Illinois    0   0.0000000
## 116           Pike   Illinois    0   0.0000000
## 117           Pope   Illinois    0   0.0000000
## 118        Pulaski   Illinois    0   0.0000000
## 119         Putnam   Illinois    0   0.0000000
## 120       Schuyler   Illinois    0   0.0000000
## 121          Scott   Illinois    0   0.0000000
## 122         Shelby   Illinois    0   8.1889715
## 123          Stark   Illinois    0   0.0000000
## 124         Wabash   Illinois    0   0.0000000
## 125         Benton    Indiana    0   0.0000000
## 126       Crawford    Indiana    0  25.7832970
## 127         Martin    Indiana    0   0.0000000
## 128         Newton    Indiana    0 162.7085900
## 129           Ohio    Indiana    0   0.0000000
## 130           Pike    Indiana    0   0.0000000
## 131    Switzerland    Indiana    0   0.0000000
## 132          Union    Indiana    0   0.0000000
## 133         Warren    Indiana    0  40.6325570
## 134          Adair       Iowa    0   0.0000000
## 135          Adams       Iowa    0   0.0000000
## 136        Audubon       Iowa    0  17.8377130
## 137        Calhoun       Iowa    0   0.0000000
## 138       Cherokee       Iowa    0   0.0000000
## 139      Chickasaw       Iowa    0   0.0000000
## 140         Clarke       Iowa    0   0.0000000
## 141          Davis       Iowa    0   0.0000000
## 142        Decatur       Iowa    0   0.0000000
## 143          Emmet       Iowa    0   0.0000000
## 144       Franklin       Iowa    0   0.0000000
## 145        Fremont       Iowa    0   0.0000000
## 146         Greene       Iowa    0   0.0000000
## 147         Grundy       Iowa    0   0.0000000
## 148        Hancock       Iowa    0   0.0000000
```

44

```
## 149          Howard         Iowa          0   0.0000000
## 150        Humboldt         Iowa          0   0.0000000
## 151             Ida         Iowa          0   0.0000000
## 152          Keokuk         Iowa          0   0.0000000
## 153          Louisa         Iowa          0  85.7722320
## 154           Lucas         Iowa          0   0.0000000
## 155            Lyon         Iowa          0   0.0000000
## 156        Mitchell         Iowa          0   0.0000000
## 157          Monona         Iowa          0   0.0000000
## 158          Monroe         Iowa          0  35.5951390
## 159      Montgomery         Iowa          0   0.0000000
## 160         Osceola         Iowa          0   0.0000000
## 161       Palo Alto         Iowa          0   0.0000000
## 162      Pocahontas         Iowa          0   0.0000000
## 163        Ringgold         Iowa          0   0.0000000
## 164             Sac         Iowa          0   0.0000000
## 165          Shelby         Iowa          0   0.0000000
## 166          Taylor         Iowa          0   0.0000000
## 167           Union         Iowa          0   0.0000000
## 168       Van Buren         Iowa          0   0.0000000
## 169           Wayne         Iowa          0   0.0000000
## 170       Winnebago         Iowa          0   0.0000000
## 171           Worth         Iowa          0   0.0000000
## 172        Anderson       Kansas          0   0.0000000
## 173          Barber       Kansas          0   0.0000000
## 174           Brown       Kansas          0   0.0000000
## 175           Chase       Kansas          0   0.0000000
## 176      Chautauqua       Kansas          0   0.0000000
## 177           Clark       Kansas          0   0.0000000
## 178            Clay       Kansas          0  22.8380530
## 179           Cloud       Kansas          0   0.0000000
## 180          Coffey       Kansas          0 236.0848500
## 181        Doniphan       Kansas          0   0.0000000
## 182         Edwards       Kansas          0   0.0000000
## 183             Elk       Kansas          0   0.0000000
## 184       Ellsworth       Kansas          0   0.0000000
## 185          Graham       Kansas          0   0.0000000
## 186           Grant       Kansas          0   0.0000000
## 187            Gray       Kansas          0   0.0000000
## 188       Greenwood       Kansas          0   0.0000000
## 189          Harper       Kansas          0   0.0000000
## 190         Haskell       Kansas          0   0.0000000
## 191          Jewell       Kansas          0   0.0000000
## 192          Kearny       Kansas          0  17.5613980
## 193         Kingman       Kansas          0   0.0000000
## 194           Kiowa       Kansas          0   0.0000000
## 195         Lincoln       Kansas          0   0.0000000
## 196            Linn       Kansas          0   0.0000000
## 197           Logan       Kansas          0   0.0000000
## 198        Marshall       Kansas          0   0.0000000
## 199           Meade       Kansas          0   0.0000000
## 200        Mitchell       Kansas          0   0.0000000
## 201          Morris       Kansas          0   0.0000000
## 202          Morton       Kansas          0   0.0000000
```

```
## 203          Nemaha      Kansas           0   0.0000000
## 204            Ness      Kansas           0   0.0000000
## 205          Norton      Kansas           0   0.0000000
## 206         Osborne      Kansas           0   0.0000000
## 207          Ottawa      Kansas           0   0.0000000
## 208          Pawnee      Kansas           0   0.0000000
## 209        Phillips      Kansas           0   0.0000000
## 210           Pratt      Kansas           0   0.0000000
## 211        Republic      Kansas           0   0.0000000
## 212            Rice      Kansas           0   0.0000000
## 213           Rooks      Kansas           0   0.0000000
## 214            Rush      Kansas           0   0.0000000
## 215         Russell      Kansas           0   0.0000000
## 216           Scott      Kansas           0   0.0000000
## 217        Sheridan      Kansas           0   0.0000000
## 218         Sherman      Kansas           0   0.0000000
## 219        Stafford      Kansas           0   0.0000000
## 220         Stevens      Kansas           0   0.0000000
## 221          Thomas      Kansas           0   0.0000000
## 222           Trego      Kansas           0   0.0000000
## 223       Wabaunsee      Kansas           0   0.0000000
## 224      Washington      Kansas           0   0.0000000
## 225          Wilson      Kansas           0   0.0000000
## 226         Woodson      Kansas           0   0.0000000
## 227         Ballard    Kentucky           0   0.0000000
## 228        Carlisle    Kentucky           0  33.5185850
## 229         Carroll    Kentucky           0   0.0000000
## 230         Clinton    Kentucky           0   0.0000000
## 231      Crittenden    Kentucky           0  25.2011590
## 232      Cumberland    Kentucky           0   0.0000000
## 233        Edmonson    Kentucky           0  87.7209240
## 234         Elliott    Kentucky           0   0.0000000
## 235          Fulton    Kentucky           0   0.0000000
## 236           Green    Kentucky           0   0.0000000
## 237         Hancock    Kentucky           0   0.0000000
## 238         Hickman    Kentucky           0   0.0000000
## 239             Lee    Kentucky           0   0.0000000
## 240      Livingston    Kentucky           0   0.0000000
## 241            Lyon    Kentucky           0 101.3395600
## 242          McLean    Kentucky           0   2.8693745
## 243          Martin    Kentucky           0   0.0000000
## 244        Metcalfe    Kentucky           0   1.1059256
## 245          Monroe    Kentucky           0   0.0000000
## 246        Nicholas    Kentucky           0   0.0000000
## 247            Owen    Kentucky           0   0.0000000
## 248          Owsley    Kentucky           0   0.0000000
## 249            Todd    Kentucky           0   0.0000000
## 250         Trimble    Kentucky           0   0.0000000
## 251      Washington    Kentucky           0   0.0000000
## 252           Wolfe    Kentucky           0   0.0000000
## 253        Caldwell   Louisiana           0   0.0000000
## 254         Cameron   Louisiana           0   0.0000000
## 255    East Carroll   Louisiana           0   0.0000000
## 256         Madison   Louisiana           0   0.0000000
```

```
## 257          Red River      Louisiana              0 156.7853500
## 258         St. Helena      Louisiana              0  23.7223450
## 259             Tensas      Louisiana              0   0.0000000
## 260       West Carroll      Louisiana              0   0.0000000
## 261          Nantucket  Massachusetts              0   0.0000000
## 262             Alcona       Michigan              0  15.9754770
## 263              Alger       Michigan              0   0.0000000
## 264             Baraga       Michigan              0   0.0000000
## 265               Lake       Michigan              0   0.0000000
## 266               Luce       Michigan              0   0.0000000
## 267           Mackinac       Michigan              0   0.0000000
## 268        Montmorency       Michigan              0   0.0000000
## 269          Ontonagon       Michigan              0   0.0000000
## 270             Oscoda       Michigan              0  14.9618470
## 271        Schoolcraft       Michigan              0   0.0000000
## 272          Big Stone      Minnesota              0   0.0000000
## 273           Chippewa      Minnesota              0   2.6715825
## 274         Clearwater      Minnesota              0   0.0000000
## 275               Cook      Minnesota              0   0.0000000
## 276         Cottonwood      Minnesota              0   0.0000000
## 277              Grant      Minnesota              0   0.0000000
## 278            Jackson      Minnesota              0   0.0000000
## 279            Kanabec      Minnesota              0   4.5390902
## 280            Kittson      Minnesota              0   0.0000000
## 281      Lac qui Parle      Minnesota              0   0.0000000
## 282               Lake      Minnesota              0   0.0000000
## 283  Lake of the Woods      Minnesota              0   0.0000000
## 284            Lincoln      Minnesota              0   0.0000000
## 285           Mahnomen      Minnesota              0  27.0239370
## 286           Marshall      Minnesota              0   0.0000000
## 287             Murray      Minnesota              0   0.0000000
## 288             Norman      Minnesota              0   0.0000000
## 289          Pipestone      Minnesota              0   0.0000000
## 290               Pope      Minnesota              0   0.0000000
## 291               Rock      Minnesota              0   0.0000000
## 292            Stevens      Minnesota              0   0.0000000
## 293              Swift      Minnesota              0   0.0000000
## 294            Wabasha      Minnesota              0   0.0000000
## 295           Watonwan      Minnesota              0   0.0000000
## 296             Wilkin      Minnesota              0 114.8437700
## 297    Yellow Medicine      Minnesota              0   0.0000000
## 298             Benton    Mississippi              0   0.0000000
## 299            Carroll    Mississippi              0 110.3927700
## 300            Choctaw    Mississippi              0  52.0725520
## 301          Claiborne    Mississippi              0  10.8561720
## 302           Franklin    Mississippi              0  32.3199540
## 303          Humphreys    Mississippi              0 125.4648100
## 304          Jefferson    Mississippi              0   2.0091753
## 305             Kemper    Mississippi              0 129.6401100
## 306         Montgomery    Mississippi              0  30.8210830
## 307            Noxubee    Mississippi              0  48.6769790
## 308              Perry    Mississippi              0  39.2060320
## 309            Quitman    Mississippi              0   0.0000000
## 310            Sharkey    Mississippi              0   0.0000000
```

```
## 311          Tunica  Mississippi           0  57.8075290
## 312         Webster  Mississippi           0  37.9976880
## 313       Wilkinson  Mississippi           0 228.2437600
## 314        Atchison     Missouri           0   0.0000000
## 315          Barton     Missouri           0   0.0000000
## 316       Bollinger     Missouri           0   0.0000000
## 317         Carroll     Missouri           0   0.0000000
## 318          Carter     Missouri           0  52.7632480
## 319        Chariton     Missouri           0   0.0000000
## 320           Clark     Missouri           0   0.0000000
## 321            Dade     Missouri           0   0.0000000
## 322         Daviess     Missouri           0   0.0000000
## 323          Gentry     Missouri           0   2.4593592
## 324          Grundy     Missouri           0   0.0000000
## 325        Harrison     Missouri           0   0.0000000
## 326         Hickory     Missouri           0   0.0000000
## 327            Holt     Missouri           0   0.0000000
## 328          Howard     Missouri           0   0.0000000
## 329            Iron     Missouri           0   0.0000000
## 330            Knox     Missouri           0   0.0000000
## 331        Lawrence     Missouri           0   0.0000000
## 332           Lewis     Missouri           0  21.6425630
## 333          Maries     Missouri           0   0.0000000
## 334          Mercer     Missouri           0   0.0000000
## 335          Monroe     Missouri           0   0.0000000
## 336      Montgomery     Missouri           0   0.0000000
## 337          Oregon     Missouri           0   0.0000000
## 338           Osage     Missouri           0   0.0000000
## 339           Ozark     Missouri           0   0.0000000
## 340          Putnam     Missouri           0   0.0000000
## 341           Ralls     Missouri           0   0.0000000
## 342        Reynolds     Missouri           0   0.0000000
## 343       St. Clair     Missouri           0   0.0000000
## 344        Schuyler     Missouri           0   0.0000000
## 345         Shannon     Missouri           0   0.0000000
## 346          Shelby     Missouri           0   0.0000000
## 347        Sullivan     Missouri           0   0.0000000
## 348           Texas     Missouri           0   0.0000000
## 349      Broadwater      Montana           0   0.0000000
## 350        Chouteau      Montana           0   0.0000000
## 351         Granite      Montana           0   0.0000000
## 352         Mineral      Montana           0   0.0000000
## 353     Musselshell      Montana           0   0.0000000
## 354        Phillips      Montana           0   0.0000000
## 355         Pondera      Montana           0   0.0000000
## 356          Powell      Montana           0   0.0000000
## 357        Sheridan      Montana           0   0.0000000
## 358     Sweet Grass      Montana           0   0.0000000
## 359           Teton      Montana           0   0.0000000
## 360           Toole      Montana           0 338.2869600
## 361        Antelope     Nebraska           0   0.0000000
## 362           Boone     Nebraska           0   0.0000000
## 363            Burt     Nebraska           0   0.0000000
## 364          Butler     Nebraska           0   0.0000000
```

```
## 365         Cherry       Nebraska          0   0.0000000
## 366           Clay       Nebraska          0   0.0000000
## 367          Deuel       Nebraska          0   0.0000000
## 368          Dixon       Nebraska          0   0.0000000
## 369       Fillmore       Nebraska          0   0.0000000
## 370       Franklin       Nebraska          0   0.0000000
## 371         Harlan       Nebraska          0   0.0000000
## 372      Jefferson       Nebraska          0   0.0000000
## 373        Johnson       Nebraska          0   0.0000000
## 374        Kearney       Nebraska          0   0.0000000
## 375          Nance       Nebraska          0   0.0000000
## 376         Nemaha       Nebraska          0   0.0000000
## 377         Pawnee       Nebraska          0   0.0000000
## 378           Polk       Nebraska          0   0.0000000
## 379     Richardson       Nebraska          0   0.0000000
## 380        Sherman       Nebraska          0   0.0000000
## 381        Stanton       Nebraska          0   0.0000000
## 382         Thayer       Nebraska          0   0.0000000
## 383       Thurston       Nebraska          0   0.0000000
## 384         Valley       Nebraska          0   0.0000000
## 385          Wayne       Nebraska          0   0.0000000
## 386        Webster       Nebraska          0   0.0000000
## 387         Lander         Nevada          0   0.0000000
## 388        Lincoln         Nevada          0   0.0000000
## 389        Mineral         Nevada          0   0.0000000
## 390       Pershing         Nevada          0   0.0000000
## 391     White Pine         Nevada          0   0.0000000
## 392         Catron     New Mexico          0  57.8115960
## 393      Guadalupe     New Mexico          0   0.0000000
## 394        Hidalgo     New Mexico          0   0.0000000
## 395           Mora     New Mexico          0   0.0000000
## 396          Union     New Mexico          0   0.0000000
## 397       Hamilton       New York          0   0.0000000
## 398      Alleghany North Carolina          0   0.0000000
## 399         Camden North Carolina          0   0.0000000
## 400           Clay North Carolina          0   0.0000000
## 401           Hyde North Carolina          0   0.0000000
## 402          Jones North Carolina          0  52.7119180
## 403        Tyrrell North Carolina          0   0.0000000
## 404         Benson   North Dakota          0   0.0000000
## 405         Bowman   North Dakota          0   0.0000000
## 406          Burke   North Dakota          0   0.0000000
## 407       Cavalier   North Dakota          0   0.0000000
## 408         Dickey   North Dakota          0   0.0000000
## 409           Dunn   North Dakota          0   0.0000000
## 410           Eddy   North Dakota          0   0.0000000
## 411        Emmons   North Dakota          0  81.3717800
## 412         Foster   North Dakota          0   0.0000000
## 413  Golden Valley   North Dakota          0   0.0000000
## 414         Griggs   North Dakota          0   0.0000000
## 415      Hettinger   North Dakota          0   0.0000000
## 416        LaMoure   North Dakota          0   0.0000000
## 417        McHenry   North Dakota          0  61.3299830
## 418       McIntosh   North Dakota          0   0.0000000
```

```
## 419          McKenzie   North Dakota             0   0.0000000
## 420            Nelson   North Dakota             0   0.0000000
## 421            Pierce   North Dakota             0   0.0000000
## 422            Ransom   North Dakota             0   0.0000000
## 423          Renville   North Dakota             0   0.0000000
## 424           Sargent   North Dakota             0   0.0000000
## 425            Traill   North Dakota             0   0.0000000
## 426             Wells   North Dakota             0   0.0000000
## 427           Clinton           Ohio             0   0.3997772
## 428           Alfalfa       Oklahoma             0   0.0000000
## 429            Beaver       Oklahoma             0   0.0000000
## 430              Coal       Oklahoma             0   0.0000000
## 431            Cotton       Oklahoma             0  55.9691540
## 432             Dewey       Oklahoma             0   0.0000000
## 433             Ellis       Oklahoma             0   0.0000000
## 434             Grant       Oklahoma             0   0.0000000
## 435             Greer       Oklahoma             0 303.6935700
## 436            Harper       Oklahoma             0   0.0000000
## 437           Haskell       Oklahoma             0   0.0000000
## 438         Jefferson       Oklahoma             0   0.0000000
## 439          Johnston       Oklahoma             0   0.0000000
## 440             Kiowa       Oklahoma             0   0.0000000
## 441           Latimer       Oklahoma             0  29.8404560
## 442              Love       Oklahoma             0   0.0000000
## 443             Major       Oklahoma             0  35.2578240
## 444             Noble       Oklahoma             0   0.0000000
## 445          Okfuskee       Oklahoma             0   0.0000000
## 446        Pushmataha       Oklahoma             0   0.0000000
## 447           Tillman       Oklahoma             0  13.3802540
## 448           Washita       Oklahoma             0   0.0000000
## 449             Woods       Oklahoma             0   0.0000000
## 450             Grant         Oregon             0   0.0000000
## 451            Harney         Oregon             0   0.0000000
## 452              Lake         Oregon             0   0.0000000
## 453            Morrow         Oregon             0   0.0000000
## 454           Wallowa         Oregon             0   0.0000000
## 455           Cameron   Pennsylvania             0   0.0000000
## 456            Forest   Pennsylvania             0   0.0000000
## 457            Fulton   Pennsylvania             0   5.7103047
## 458          Sullivan   Pennsylvania             0   0.0000000
## 459         Allendale South Carolina             0  47.4861370
## 460         McCormick South Carolina             0  30.4945660
## 461            Aurora   South Dakota             0   0.0000000
## 462             Clark   South Dakota             0   0.0000000
## 463           Edmunds   South Dakota             0   0.0000000
## 464             Faulk   South Dakota             0   0.0000000
## 465            Hanson   South Dakota             0   0.0000000
## 466           Jerauld   South Dakota             0 100.5503800
## 467             Miner   South Dakota             0   0.0000000
## 468           Perkins   South Dakota             0   0.0000000
## 469            Potter   South Dakota             0   0.0000000
## 470           Sanborn   South Dakota             0   0.0000000
## 471            Cannon      Tennessee             0   0.0000000
## 472              Clay      Tennessee             0   0.0000000
```

```
## 473           Decatur     Tennessee          0   0.0000000
## 474          Grainger     Tennessee          0   0.0000000
## 475           Hancock     Tennessee          0   0.0000000
## 476           Houston     Tennessee          0   0.0000000
## 477           Jackson     Tennessee          0   0.0000000
## 478              Lake     Tennessee          0   0.0000000
## 479             Lewis     Tennessee          0   0.0000000
## 480             Meigs     Tennessee          0   0.0000000
## 481             Moore     Tennessee          0   0.0000000
## 482             Perry     Tennessee          0   0.0000000
## 483           Pickett     Tennessee          0   0.0000000
## 484         Trousdale     Tennessee          0  90.9448090
## 485         Van Buren     Tennessee          0   0.0000000
## 486           Andrews         Texas          0   0.0000000
## 487            Archer         Texas          0   0.0000000
## 488         Armstrong         Texas          0   0.0000000
## 489            Bailey         Texas          0   0.0000000
## 490            Baylor         Texas          0   0.0000000
## 491            Blanco         Texas          0   1.1234882
## 492            Bosque         Texas          0   0.0000000
## 493          Brewster         Texas          0   0.0000000
## 494            Brooks         Texas          0   0.0000000
## 495              Camp         Texas          0   0.0000000
## 496            Carson         Texas          0   0.0000000
## 497            Castro         Texas          0  40.4240340
## 498         Childress         Texas          0   0.0000000
## 499              Clay         Texas          0   0.0000000
## 500           Cochran         Texas          0   0.0000000
## 501              Coke         Texas          0   0.0000000
## 502           Coleman         Texas          0   0.0000000
## 503     Collingsworth         Texas          0   0.0000000
## 504            Concho         Texas          0   0.0000000
## 505             Crane         Texas          0  11.7442870
## 506          Crockett         Texas          0   0.0000000
## 507            Crosby         Texas          0  40.9019320
## 508         Culberson         Texas          0   0.0000000
## 509             Delta         Texas          0   0.0000000
## 510           Dickens         Texas          0   0.0000000
## 511            Dimmit         Texas          0   0.0000000
## 512            Donley         Texas          0   0.0000000
## 513             Duval         Texas          0   0.0000000
## 514            Fisher         Texas          0  12.3194860
## 515             Floyd         Texas          0   0.0000000
## 516          Franklin         Texas          0   0.0000000
## 517             Garza         Texas          0   0.0000000
## 518            Goliad         Texas          0   0.0000000
## 519              Hall         Texas          0   0.0000000
## 520          Hamilton         Texas          0   6.0317268
## 521          Hansford         Texas          0  56.3091890
## 522          Hardeman         Texas          0   0.0000000
## 523           Haskell         Texas          0   0.0000000
## 524          Hemphill         Texas          0   0.0000000
## 525              Jack         Texas          0   0.0000000
## 526          Jim Hogg         Texas          0   0.0000000
```

```
## 527           Jones        Texas      0  0.0000000
## 528         Kimble        Texas      0  0.0000000
## 529         Kinney        Texas      0  0.0000000
## 530           Knox        Texas      0  0.0000000
## 531       La Salle        Texas      0  0.0000000
## 532       Lipscomb        Texas      0  0.0000000
## 533       Live Oak        Texas      0  0.0000000
## 534           Lynn        Texas      0 37.5891610
## 535      McCulloch        Texas      0  0.0000000
## 536        Madison        Texas      0  0.0000000
## 537         Martin        Texas      0 30.8631380
## 538         Menard        Texas      0  0.0000000
## 539          Mills        Texas      0  0.0000000
## 540       Mitchell        Texas      0  0.0000000
## 541         Morris        Texas      0  0.0000000
## 542         Newton        Texas      0  0.0000000
## 543      Ochiltree        Texas      0 33.3565140
## 544         Parmer        Texas      0  0.0000000
## 545       Presidio        Texas      0  0.0000000
## 546          Rains        Texas      0  0.0000000
## 547           Real        Texas      0  0.0000000
## 548         Reeves        Texas      0  0.0000000
## 549        Refugio        Texas      0  0.0000000
## 550       Rockwall        Texas      0 16.5594480
## 551        Runnels        Texas      0  0.0000000
## 552         Sabine        Texas      0  0.0000000
## 553  San Augustine        Texas      0 40.1331900
## 554     San Jacinto        Texas      0  0.0000000
## 555       San Saba        Texas      0  0.0000000
## 556     Schleicher        Texas      0  0.0000000
## 557    Shackelford        Texas      0  0.0000000
## 558        Sherman        Texas      0  0.0000000
## 559      Somervell        Texas      0  0.0000000
## 560       Stephens        Texas      0  0.0000000
## 561         Sutton        Texas      0  0.0000000
## 562        Swisher        Texas      0  0.0000000
## 563          Terry        Texas      0  0.0000000
## 564          Upton        Texas      0  0.0000000
## 565           Ward        Texas      0  0.0000000
## 566        Wheeler        Texas      0  0.0000000
## 567      Wilbarger        Texas      0  0.0000000
## 568        Winkler        Texas      0  0.0000000
## 569         Yoakum        Texas      0  0.0000000
## 570         Zapata        Texas      0  0.0000000
## 571         Zavala        Texas      0  0.0000000
## 572         Beaver         Utah      0  0.0000000
## 573       Garfield         Utah      0  0.0000000
## 574           Kane         Utah      0  0.0000000
## 575       San Juan         Utah      0 52.1456410
## 576         Amelia     Virginia      0 18.6347540
## 577           Bath     Virginia      0  0.0000000
## 578          Bland     Virginia      0  0.0000000
## 579   Charles City     Virginia      0 32.1062470
## 580         Clarke     Virginia      0  0.0000000
```

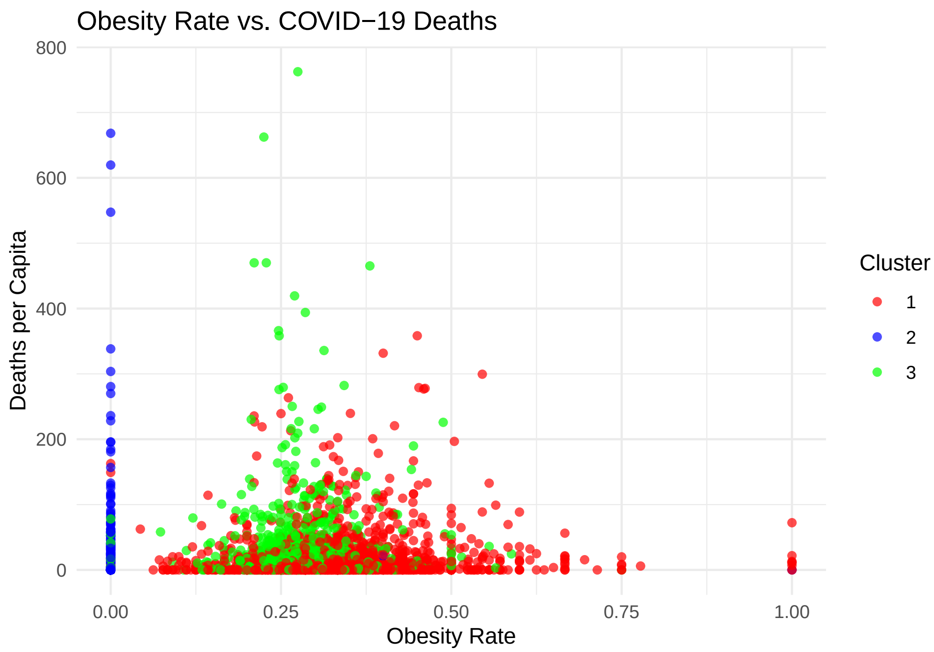
```
## 581               Craig       Virginia            0   0.0000000
## 582          Cumberland       Virginia            0   0.0000000
## 583               Essex       Virginia            0   0.0000000
## 584           Goochland       Virginia            0  55.0697360
## 585      King and Queen       Virginia            0   3.2287488
## 586         King George       Virginia            0  41.4670100
## 587           Lancaster       Virginia            0   0.0000000
## 588             Mathews       Virginia            0   0.0000000
## 589           Middlesex       Virginia            0   0.0000000
## 590              Nelson       Virginia            0   0.0000000
## 591        Rappahannock       Virginia            0   0.0000000
## 592            Richmond       Virginia            0  57.5768700
## 593               Surry       Virginia            0  33.1339910
## 594        Bristol city       Virginia            0   0.0000000
## 595    Buena Vista city       Virginia            0   0.0000000
## 596        Emporia city       Virginia            0 117.0800200
## 597   Falls Church city       Virginia            0  77.9791560
## 598 Fredericksburg city       Virginia            0   0.0000000
## 599          Galax city       Virginia            0   0.0000000
## 600         Norton city       Virginia            0   0.0000000
## 601       Poquoson city       Virginia            0   0.0000000
## 602       Staunton city       Virginia            0   0.0000000
## 603     Waynesboro city       Virginia            0   0.0000000
## 604     Winchester city       Virginia            0   4.9519019
## 605             Calhoun  West Virginia            0   0.0000000
## 606                Clay  West Virginia            0   0.0000000
## 607           Doddridge  West Virginia            0   0.0000000
## 608              Gilmer  West Virginia            0   0.0000000
## 609               Grant  West Virginia            0   0.0000000
## 610           Pendleton  West Virginia            0   0.0000000
## 611           Pleasants  West Virginia            0   0.0000000
## 612          Pocahontas  West Virginia            0   0.0000000
## 613             Ritchie  West Virginia            0   0.0000000
## 614              Tucker  West Virginia            0   0.0000000
## 615               Tyler  West Virginia            0   0.0000000
## 616             Webster  West Virginia            0   0.0000000
## 617                Wirt  West Virginia            0   0.0000000
## 618            Florence      Wisconsin            0   0.0000000
## 619              Forest      Wisconsin            0   3.4906044
## 620                Iron      Wisconsin            0  58.7476080
## 621           Menominee      Wisconsin            0   0.0000000
## 622               Pepin      Wisconsin            0   0.0000000
```

```r
sum(is.na(data$bmi_obese_q1))
```

```
## [1] 0
```

```r
selected_features <- c("bmi_obese_q1", "deathspc")
df_cluster <- data_cleaned %>%
  select(all_of(selected_features)) %>%
  na.omit()
df_cluster$cluster <- as.factor(kmeans_result$cluster)
df_cluster$cluster <- as.factor(kmeans_result$cluster)
plot_obesity <- ggplot(df_cluster, aes(x = bmi_obese_q1, y = deathspc, color = cluster)) +
  geom_point(alpha = 0.7) +
```

```
  scale_color_manual(values = c("red", "blue", "green")) +
  labs(title = "Obesity Rate vs. COVID-19 Deaths",
       x = "Obesity Rate", y = "Deaths per Capita", color = "Cluster") +
  theme_minimal()

print(plot_obesity)
```



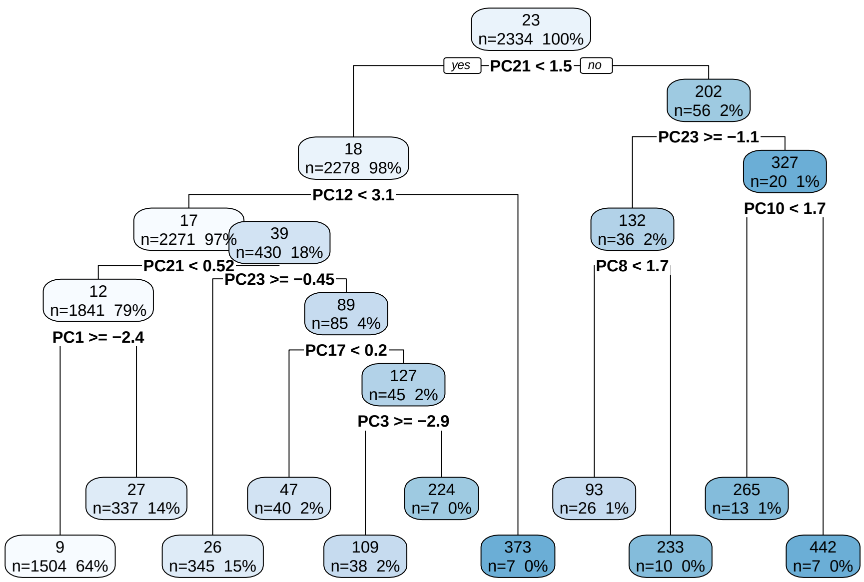

```
ggsave("Obesity_vs_DeathRate.png", plot = plot_obesity, width = 8, height = 6, dpi = 300)
```

Interpretation of Obesity Rate vs. COVID-19 Mortality Clustering: This scatter plot visualizes the relationship between obesity rates and COVID-19 deaths per capita, with counties grouped into three clusters using K-Means clustering. The X-axis represents obesity rates, while the Y-axis represents COVID-19 deaths per capita. The different colors represent the clusters: Cluster 1 (Red): Counties with higher COVID-19 mortality, primarily falling within moderate obesity rates (~0.10 - 0.50). Cluster 2 (Blue): Counties with low mortality but concentrated at an obesity rate of 0.00. This suggests these counties either have missing or inaccurately reported obesity data. Cluster 3 (Green): Counties with lower mortality rates, spread across various obesity levels, including some with high obesity rates.

Key Insights: 1. Obesity is correlated with higher COVID-19 mortality, but not exclusively. Cluster 1 (Red, high-mortality counties) consists mostly of counties with moderate obesity rates (0.10 - 0.50). This supports medical research that obesity increases the risk of severe COVID-19 outcomes, but other factors like healthcare access, socio-economic conditions, and chronic illnesses may also be influencing mortality. 2. Cluster 2 (Blue) counties with 0.00 obesity rates still exist but with low mortality. Despite replacing 0.00 values with NA and re-running the clustering, these counties still appear as a distinct group. This suggests that these counties may not have reported obesity data correctly or have exceptionally low obesity rates. 3. Counties with high obesity rates (~0.50+) are mostly in Cluster 3 (Green) with lower mortality. Some high-obesity counties do not have high COVID-19 deaths, which could indicate better healthcare systems, stronger public health policies, or lower population density as protective factors. 4. Some low-obesity counties still experience high mortality. A few Cluster 1 (Red) counties with obesity rates under 0.25 still have high deaths per capita. This suggests that other factors, such as diabetes prevalence, air pollution (PM2.5), or healthcare disparities, may contribute to high COVID-19 mortality.

##Test for nonlinearity ### Decision Tree regressor

```r
# Train a Decision Tree
tree_model <- rpart(deathspc ~ ., data = train_data, method = "anova")
# Plot the tree
rpart.plot(tree_model, type = 2, extra = 101, tweak = 1.2)
```




```r
# Get feature importance
tree_importance <- tree_model$variable.importance
print(tree_importance) # Print importance scores
```

```
##        PC21         PC12         PC23          PC8         PC13          PC3
## 2082012.8333 1053327.2732  837747.3302  641453.9050  614352.6808  612719.5236
##          PC9         PC17         PC42         PC28          PC7          PC1
##  607756.8069  379959.7247  378164.0186  317787.2713  170944.5630  160369.7249
##         PC10          PC4          PC5         PC11         PC16          PC2
##  148472.9273  117544.2387   91330.5110   77480.5967   65605.7050   56024.4932
##         PC18         PC38         PC27         PC22         PC19          PC6
##   50907.4523   50907.4523   44481.0729   22240.5365    9462.2573    8017.8443
##         PC15         PC41         PC45         PC51
##    2849.4546    2279.5636    1945.2131     833.6627
```

##Random Forest

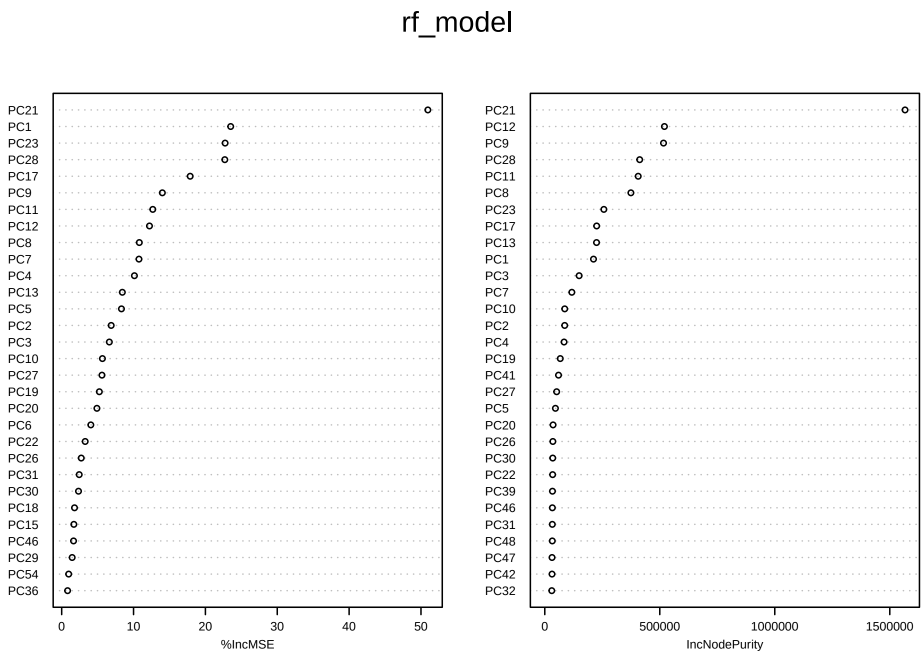
```r
set.seed(421)
# Train a Random Forest model
rf_model <- randomForest(deathspc ~ ., data = train_data, ntree = 500, importance = TRUE)
```

```r
# View feature importance
importance(rf_model)
```

```
##          %IncMSE IncNodePurity
## PC1  23.51718909     211853.00
## PC2   6.88738864      86603.58
## PC3   6.62992323     149410.90
## PC4  10.12207937      83807.36
## PC5   8.31781709      46276.29
## PC6   4.05493244      28962.88
## PC7  10.75469296     117639.42
## PC8  10.80209804     374405.23
## PC9  14.00655333     516316.37
## PC10  5.66828730      86707.82
## PC11 12.67692578     406165.61
## PC12 12.21629083     520190.83
## PC13  8.43885298     225084.72
## PC14  0.22172323      28760.03
## PC15  1.69201122      28670.30
## PC16  0.23663810      29789.50
## PC17 17.88152500     225594.75
## PC18  1.78515268      27085.47
## PC19  5.23939868      67361.89
## PC20  4.90133223      36054.70
## PC21 50.95851092    1566356.64
## PC22  3.26872361      34166.16
## PC23 22.74187614     256873.61
## PC24 -0.17654962      21012.59
## PC25  0.18065821      20187.51
## PC26  2.72079952      35133.50
## PC27  5.60526456      51665.84
## PC28 22.70072719     412630.49
## PC29  1.45637877      26894.72
## PC30  2.34141502      34642.53
## PC31  2.43586548      32760.59
## PC32  0.39641308      30396.96
## PC33 -0.18294215      28542.21
## PC34  0.37098664      26146.88
## PC35  0.36434006      28869.67
## PC36  0.82073224      21928.14
## PC37  0.08368500      26612.61
## PC38  0.74388869      19269.46
## PC39  0.30630682      33411.66
## PC40 -0.76255312      19910.53
## PC41  0.54941081      59779.71
## PC42  0.68787197      31505.63
## PC43 -0.03966274      25387.52
## PC44  0.80316483      21745.20
## PC45 -1.17504784      19662.83
## PC46  1.64747677      32977.25
## PC47 -1.10915207      31610.08
## PC48 -0.85670789      32682.68
## PC49 -2.11232330      20742.80
## PC50 -0.28019817      24069.21
```

```
## PC51  0.18272199      19301.56
## PC52  0.23754935      21444.89
## PC53  0.08852589      22884.44
## PC54  0.97086789      25797.77
```

```r
varImpPlot(rf_model, cex = 0.5)
```




```r
# Get importance scores
importance_scores <- importance(rf_model)
# Print importance scores
print(importance_scores)
```

```
##          %IncMSE IncNodePurity
## PC1   23.51718909     211853.00
## PC2    6.88738864      86603.58
## PC3    6.62992323     149410.90
## PC4   10.12207937      83807.36
## PC5    8.31781709      46276.29
## PC6    4.05493244      28962.88
## PC7   10.75469296     117639.42
## PC8   10.80209804     374405.23
## PC9   14.00655333     516316.37
## PC10   5.66828730      86707.82
## PC11  12.67692578     406165.61
## PC12  12.21629083     520190.83
## PC13   8.43885298     225084.72
## PC14   0.22172323      28760.03
## PC15   1.69201122      28670.30
```

```
## PC16  0.23663810      29789.50
## PC17 17.88152500     225594.75
## PC18  1.78515268      27085.47
## PC19  5.23939868      67361.89
## PC20  4.90133223      36054.70
## PC21 50.95851092    1566356.64
## PC22  3.26872361      34166.16
## PC23 22.74187614     256873.61
## PC24 -0.17654962      21012.59
## PC25  0.18065821      20187.51
## PC26  2.72079952      35133.50
## PC27  5.60526456      51665.84
## PC28 22.70072719     412630.49
## PC29  1.45637877      26894.72
## PC30  2.34141502      34642.53
## PC31  2.43586548      32760.59
## PC32  0.39641308      30396.96
## PC33 -0.18294215      28542.21
## PC34  0.37098664      26146.88
## PC35  0.36434006      28869.67
## PC36  0.82073224      21928.14
## PC37  0.08368500      26612.61
## PC38  0.74388869      19269.46
## PC39  0.30630682      33411.66
## PC40 -0.76255312      19910.53
## PC41  0.54941081      59779.71
## PC42  0.68787197      31505.63
## PC43 -0.03966274      25387.52
## PC44  0.80316483      21745.20
## PC45 -1.17504784      19662.83
## PC46  1.64747677      32977.25
## PC47 -1.10915207      31610.08
## PC48 -0.85670789      32682.68
## PC49 -2.11232330      20742.80
## PC50 -0.28019817      24069.21
## PC51  0.18272199      19301.56
## PC52  0.23754935      21444.89
## PC53  0.08852589      22884.44
## PC54  0.97086789      25797.77
```

```r
# Tune Random Forest
tuned_rf <- randomForest(deathspc ~ ., data = train_data, ntree = 1000, mtry = 5, nodesize = 5)
# Predict on test set
rf_predictions <- predict(tuned_rf, newdata = test_data)
# Compute RMSE
rf_rmse <- sqrt(mean((rf_predictions - test_data$deathspc)^2))
cat("Random Forest RMSE:", rf_rmse, "\n")
```

```
## Random Forest RMSE: 27.99714
```