

Longevity in the US 2019
Continental US States

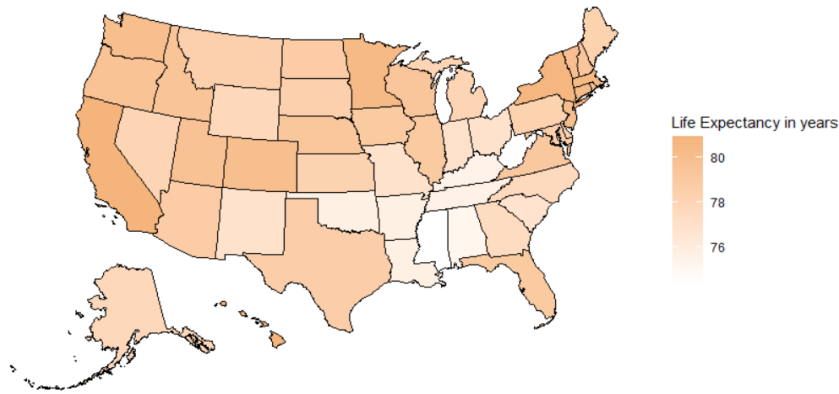


Figure 2: Heat Map of Avg. Longevity in the United States (2019)

The heat maps displayed in Figure 1 & Figure 2 present an insightful overview of life expectancy trends across the United States during two distinct periods: 2016 - 2018 and 2019. The upper map shows the average life expectancy by state for 2016 - 2018, while the lower map depicts the same metric for 2019. Notably, the states with the highest average life expectancy remain consistent across both time frames, with California, Minnesota, and New York leading in longevity. Conversely, Mississippi, West Virginia, and Alabama consistently exhibit the lowest life expectancy during these periods. This consistency between the training data (2016 - 2018) and the validation data (2019) bodes well for our model, suggesting that the underlying patterns and relationships in the data are stable over time.

Distribution of Predictors

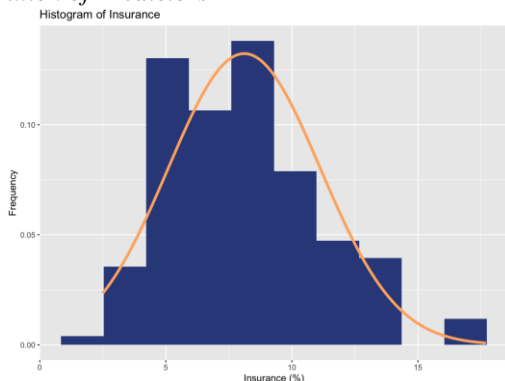


Figure 3: Histogram of Variable: 'Insurance' (2016 - 2019)

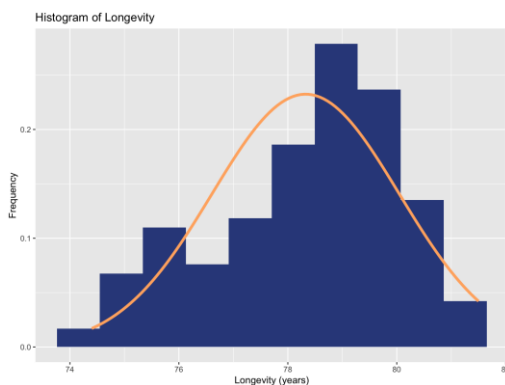


Figure 4: Histogram of Variable: 'Longevity' (2016 - 2019)

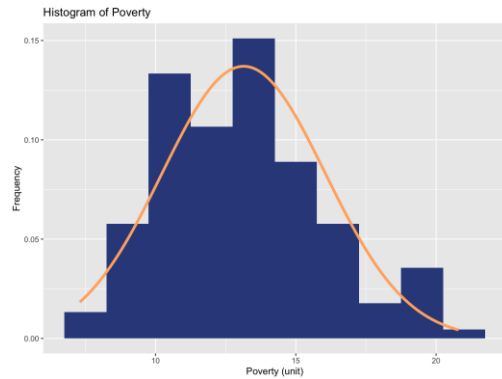


Figure 5: Histogram of Variable: 'Poverty' (2016 - 2019)

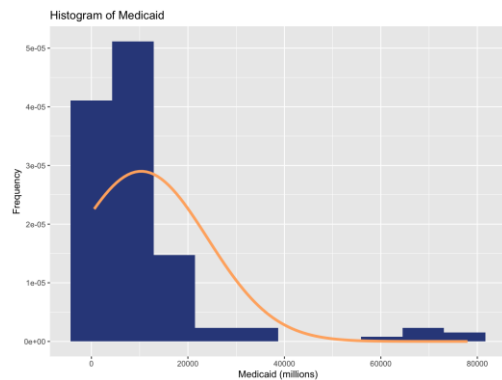


Figure 6: Histogram of Variable: 'Medicaid' (2016 - 2019)

Figures 3 - 6 delve into the distribution of key variables used in our analysis, emphasizing the importance of understanding their underlying characteristics. We focused on four crucial predictors: the percentage of uninsured population, poverty rate, life expectancy, and Medicaid estimates by state residence. Histograms illustrate that most variables, including the uninsured population percentage, poverty rate, and life expectancy, exhibit a normal distribution, indicating that the data is symmetrically distributed around the mean. However, Figure 6, which represents Medicaid estimates, reveals a right - skewed distribution with outliers. This deviation from normality in the Medicaid estimates underscores the need for careful consideration when interpreting its influence on healthcare outcomes, as skewed distributions can introduce bias and affect model performance. Understanding these distributions is pivotal for subsequent modeling steps,

informing decisions on data transformations and the selection of statistical methods. *Scatter Plots Analysis*

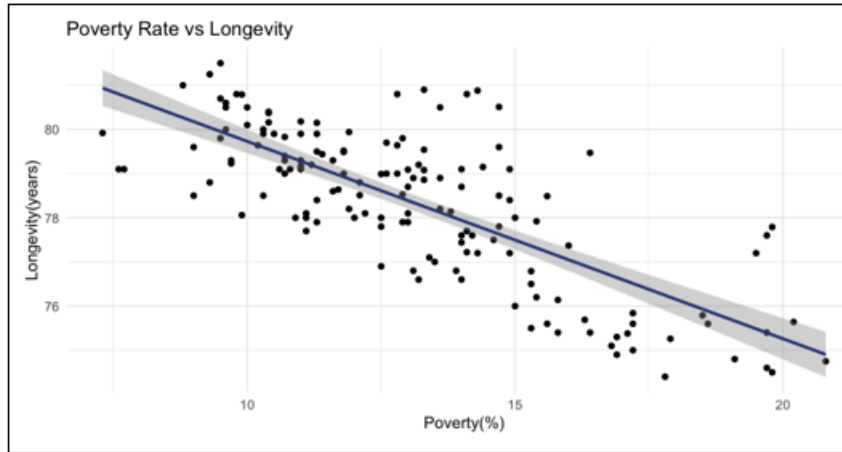


Figure 7: Scatter Plot: Poverty Rate vs. Longevity (2016 - 2019)

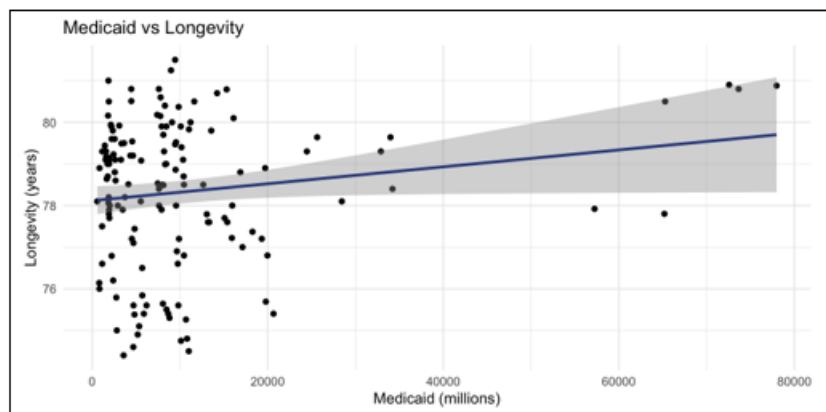


Figure 8: Scatter Plot: Medicaid vs. Longevity (2016 - 2019)

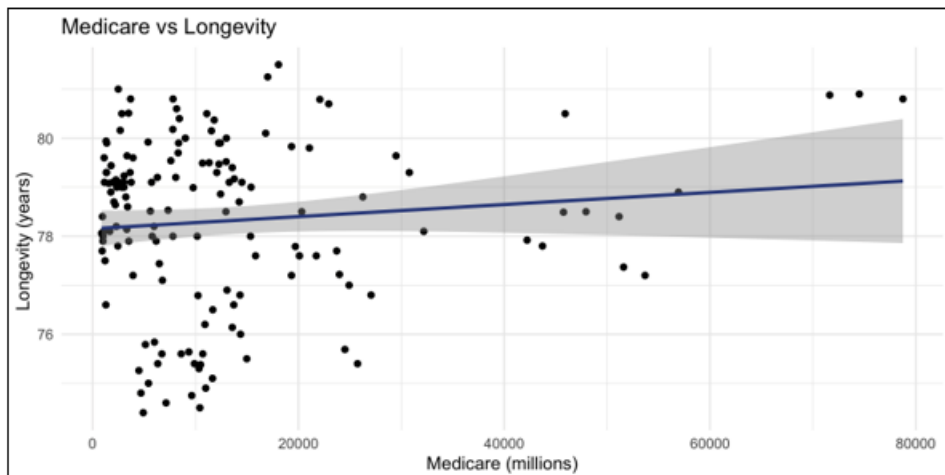


Figure 9: Scatter Plot: Medicare vs. Longevity (2016 - 2019)

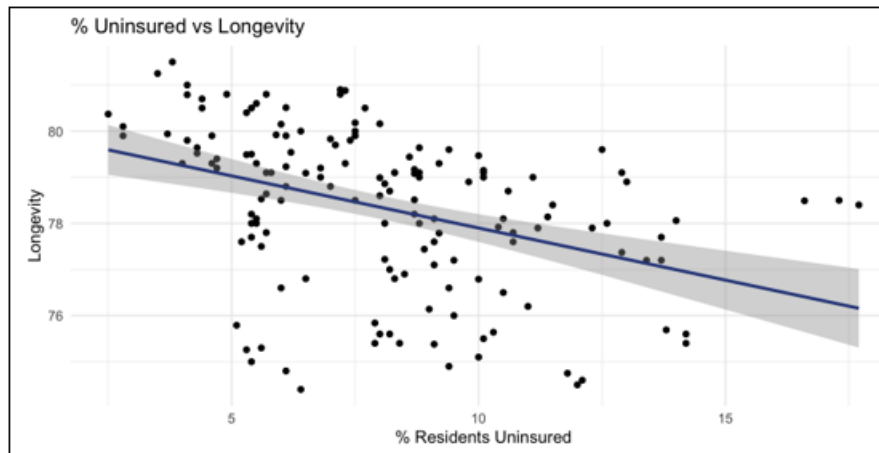


Figure 10: Scatter Plot: % Uninsured vs. Longevity (2016 - 2019)

The scatter plots in Figures 7 - 10 were generated to examine the relationships between life expectancy and several key independent variables—poverty rate, Medicaid expenditures, Medicare estimates, and the percentage of uninsured residents—across the United States from 2016 to 2019. By visually assessing these plots, we aimed to identify any strong linear correlations that might indicate these variables' predictive power regarding longevity.

Figure 7 illustrates the relationship between poverty rate and life expectancy, which shows a moderate negative correlation, with an R - squared value of 0.574 and a highly significant p - value ($p < 2e - 16$). This suggests that as the poverty rate increases, the average life expectancy decreases, a finding that aligns with existing literature and highlights poverty as a significant factor affecting health outcomes.

In contrast, Figure 8, which examines Medicaid expenditures per state, shows a weaker positive correlation with life expectancy (R - squared = 0.0265, p - value = 0.047). Although statistically significant, the low R - squared value suggests that Medicaid expenditures alone may not strongly predict longevity, indicating that other factors likely play a more critical role.

Similarly, Figure 9 displays the scatter plot for Medicare estimates against life expectancy, which reveals an even weaker correlation (R - squared = 0.0102, p - value = 0.22). The near - zero R - squared value and non - significant p - value imply that Medicare estimates do not strongly correlate with life expectancy at the state level, further suggesting that Medicare's impact on longevity might be overshadowed by other variables.

Finally, Figure 10 examines the percentage of uninsured residents against life expectancy. While this plot demonstrates a negative correlation (R - squared = 0.157, p - value = $5e - 07$), the relationship is modest. This suggests that while higher rates of uninsured individuals might contribute to lower life expectancy, the impact is moderate and likely interacts with other social and economic factors.

Overall, the scatter plots indicate that none of the examined variables alone have a sufficiently strong relationship with longevity, leading us to conclude that a more comprehensive approach, such as multiple regression modeling, is necessary to better capture the complex interplay of factors influencing life expectancy.

Models

Multiple Regression Model

	Min	1Q	Median	3Q	Max
	-1.617	-0.586	-0.026	0.620	1.834

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.62e+03	3.47e+02	4.66	7.3e-06	***
year	-7.59e-01	1.72e-01	-4.41	2.1e-05	***
uninsured	-5.19e-02	2.84e-02	-1.83	0.06971	.
Medicare	2.95e-08	8.78e-06	0.00	0.99733	
Medicaid	1.10e-05	8.63e-06	1.27	0.20514	
cost.pc	8.38e-05	2.94e-05	2.85	0.00502	**
GDP	7.68e-13	2.04e-13	3.76	0.00025	***
Unemployment	-3.98e-01	1.05e-01	-3.78	0.00023	***
poverty	-4.16e-01	7.05e-02	-5.90	2.7e-08	***
Grad.Rate	-1.26e-01	1.93e-02	-6.54	1.1e-09	***
income	1.97e-05	1.75e-05	1.13	0.26057	
gini	1.64e+01	5.38e+00	3.04	0.00286	**
burden	-4.00e+00	2.93e+00	-1.36	0.17513	

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.785 on 137 degrees of freedom
 Multiple R-squared: 0.808, Adjusted R-squared: 0.791
 F-statistic: 47.9 on 12 and 137 DF, p-value: <2e-16

Figure 11: Summary Statistics of Initial Multiple Regression Model

To assess the relationship between longevity and various predictors, we employed a multiple linear regression model that incorporated all the variables listed in our dataset. The initial results, as displayed in the regression output (Figure 11), highlight which factors are statistically significant in predicting life expectancy across different states.

Our regression analysis indicates that several variables are significant predictors of longevity, as evidenced by their p-values ($p < 0.05$). These include the year, Medicaid, cost per

capita (cost. pc), GDP, unemployment, poverty rate, graduation rate, income, and Gini index. The model yields an adjusted R-squared value of 0.791, suggesting that approximately 79.1% of the variance in life expectancy can be explained by these variables collectively. This robust R-squared value underscores the model's effectiveness in capturing the complex interactions that influence life expectancy.

Assumptions Check

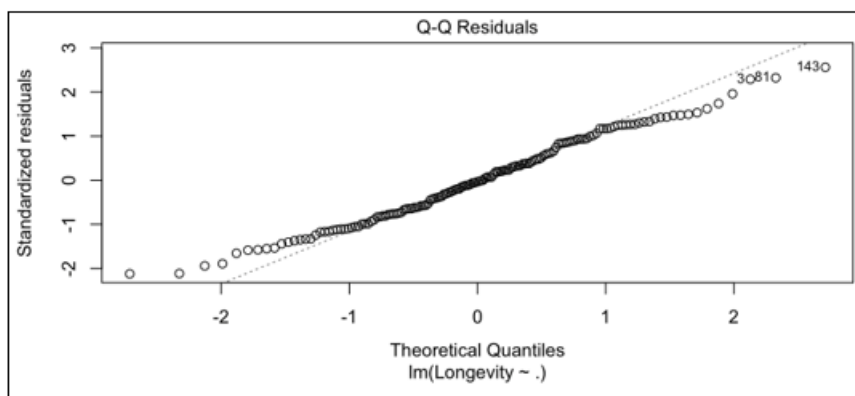


Figure 12: Quantile - Quantile Plot of Residuals (of Longevity)

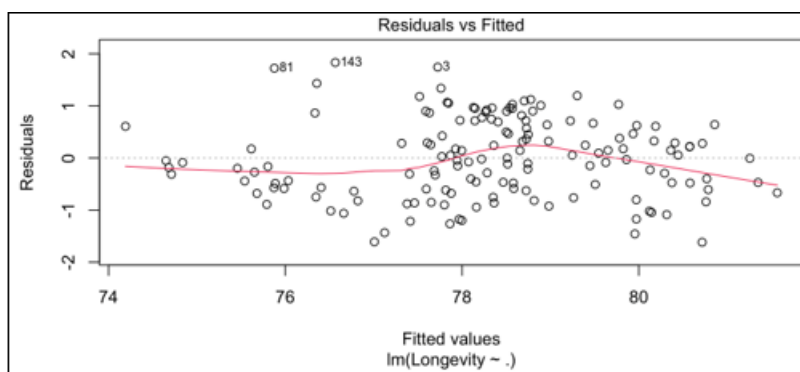


Figure 13: Residuals vs Fitted Plot (Longevity)

To ensure the validity of our regression model, we examined the three key assumptions of linear regression: normality, homoscedasticity, and linearity. The quantile - quantile (Q - Q) plot of the residuals (Figure 12) indicates that the normality assumption is satisfied, as the residuals closely follow the 45 - degree line, with only minor deviations at the extremes. For homoscedasticity, the residuals vs. fitted values plot (Figure 13) shows that the variance of the residuals is constant, with no visible pattern in their spread. The same plot

also confirms the linearity assumption, as there is no discernible curvature, indicating a linear relationship between the predictors and the dependent variable.

Overall, the diagnostic plots confirm that our regression model meets the necessary assumptions, lending credibility to the results and supporting the use of this model for further analysis.

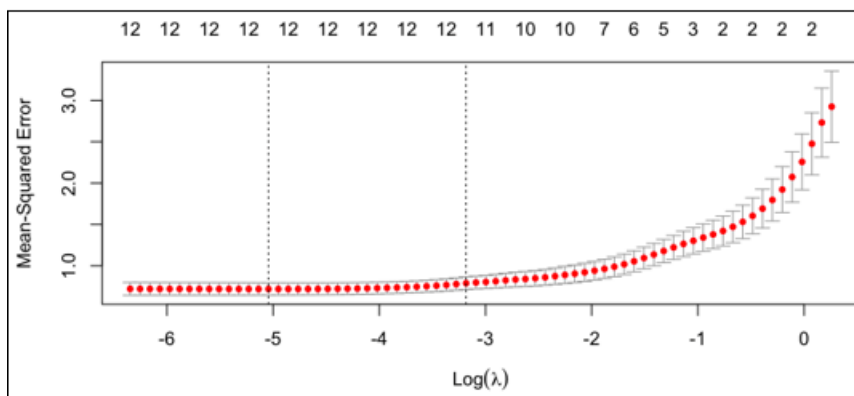


Figure 14: Mean - Squared Error vs Log (λ)

To refine our regression model, we employed the Least Absolute Shrinkage and Selection Operator (LASSO) technique, visualized in Figure 14. The graph depicts the relationship between the mean squared error (MSE) and the logarithm of the regularization parameter λ. From the plot, we determined that the minimum MSE corresponds to a log (λ) value of approximately - 5, which translates to a λ value of around 0.006.

The λ parameter in LASSO is a crucial component that controls the strength of the regularization applied to the model. Regularization is a technique used to prevent overfitting by penalizing the absolute size of the regression coefficients. The LASSO method specifically drives some coefficients to zero, effectively selecting a simpler model by eliminating less significant predictors.

Interestingly, LASSO did not eliminate any variables from our model. This outcome can be attributed to using the cv.glmnet function, which performs cross - validation and may yield different results than using the glmnet function alone. Although LASSO typically simplifies the model by removing less significant variables, in this instance, it retained all the predictors, suggesting that they all contribute to the model in some capacity.

A smaller λ value, such as 0.006 in our case, implies that the regularization effect is relatively weak, allowing more predictors to remain in the model. Conversely, a larger λ would increase the penalty on the coefficients, potentially shrinking more of them to zero and simplifying the model further. The choice of λ balances the trade - off between model complexity and prediction accuracy.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.62e+03	3.47e+02	4.66	7.3e-06	***
year	-7.59e-01	1.72e-01	-4.41	2.1e-05	***
uninsured	-5.19e-02	2.84e-02	-1.83	0.06971	.
Medicare	2.95e-08	8.78e-06	0.00	0.99733	
Medicaid	1.10e-05	8.63e-06	1.27	0.20514	
cost.pc	8.38e-05	2.94e-05	2.85	0.00502	**
GDP	7.68e-13	2.04e-13	3.76	0.00025	***
Unemployment	-3.98e-01	1.05e-01	-3.78	0.00023	***
poverty	-4.16e-01	7.05e-02	-5.90	2.7e-08	***
Grad.Rate	-1.26e-01	1.93e-02	-6.54	1.1e-09	***
income	1.97e-05	1.75e-05	1.13	0.26057	
gini	1.64e+01	5.38e+00	3.04	0.00286	**
burden	-4.00e+00	2.93e+00	-1.36	0.17513	

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 15: Summary Statistics after Zero Backwards Elimination Iteration

Given that LASSO did not reduce the number of variables, we decided to resort to backward selection for variable elimination to improve the model's statistical significance. In our first iteration of backward selection, we identified Medicare as the variable with the highest p - value (0.9973), as shown in Figure 15. At the 0.05 significance level, this p -

value indicates that Medicare does not contribute meaningfully to the model. Therefore, we eliminated Medicare from the model and reran the regression to reassess the summary statistics, focusing on the p - values of the remaining variables to determine further steps for model refinement.

Anova Table (Type II tests)

Response: Longevity					
	Sum Sq	Df	F	value	Pr(>F)
year	12.1	1	19.82	1.7e-05	***
uninsured	2.5	1	4.06	0.04593	*
Medicaid	2.8	1	4.56	0.03442	*
cost.pc	5.0	1	8.24	0.00474	**
GDP	8.9	1	14.50	0.00021	***
Unemployment	9.1	1	14.84	0.00018	***
poverty	21.5	1	35.13	2.3e-08	***
Grad.Rate	26.6	1	43.49	8.4e-10	***
income	0.8	1	1.29	0.25862	
gini	6.0	1	9.80	0.00213	**
burden	1.2	1	1.97	0.16246	
Residuals	84.4	138			

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 16: Summary Statistics after One Backwards Elimination Iterations

Continuing with our backward selection process, we observed that after removing the Medicare variable, there were still statistically insignificant p - values present in the model, as shown in Figure 16. The variable with the highest p - value

was now 'income,' with a p - value of 0.25862. Given that this exceeds our significance level of 0.05, we eliminated the income variable and reran the regression model, marking this as our second iteration.

Anova Table (Type II tests)

Response: Longevity					
	Sum Sq	Df	F	value	Pr(>F)
year	11.9	1	19.44	2.1e-05	***
uninsured	2.4	1	3.88	0.05089	.
Medicaid	3.3	1	5.31	0.02273	*
cost.pc	5.3	1	8.66	0.00381	**
GDP	9.4	1	15.36	0.00014	***
Unemployment	8.3	1	13.57	0.00033	***
poverty	130.1	1	212.26	< 2e-16	***
Grad.Rate	27.3	1	44.59	5.4e-10	***
gini	8.4	1	13.67	0.00031	***
burden	2.4	1	3.92	0.04975	*
Residuals	85.2	139			

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 17: Summary Statistics after Two Backwards Elimination Iterations

After the removal of income, the highest remaining p - value belonged to the variable 'uninsured,' which had a p - value of 0.05089, as shown in Figure 17. Though this p - value is just above our significance threshold, it still indicates that 'uninsured' is not statistically significant at the 0.05 level. Consequently, we removed the 'uninsured' variable and reran the regression model.

With the removal of 'Medicare,' 'income,' and 'uninsured,' our final multiple regression model showed that all remaining variables had p - values below the 0.05 significance level. This refinement process successfully led us to a statistically significant model, ensuring that all included variables contribute meaningfully to predicting longevity.

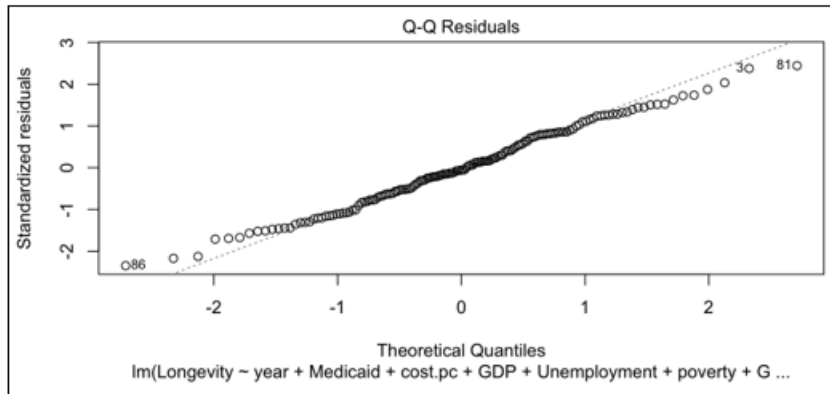


Figure 18: Summary Statistics after Three Backwards Elimination Iterations

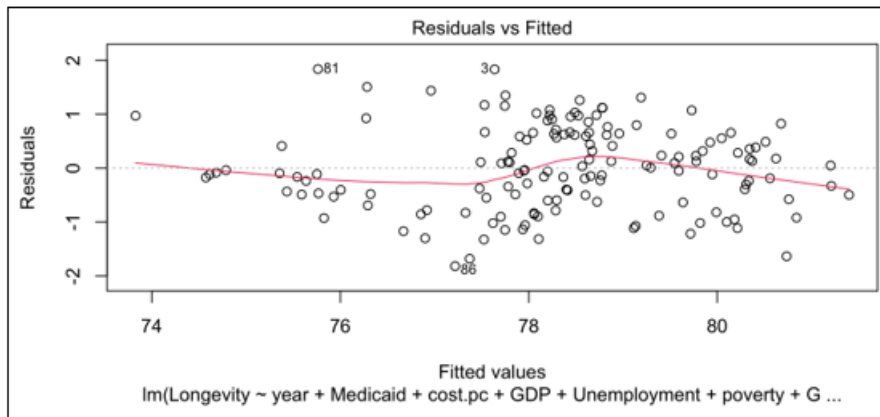


Figure 19: Summary Statistics after Three Backwards Elimination Iterations

To ensure the validity of this final regression model, we revisited the model diagnostics. The assumptions of normality, homoscedasticity, and linearity were checked once again with the new figures. The quantile - quantile (Q - Q) plot of the residuals still indicated that the normality assumption was satisfied, as the residuals closely followed the

45 - degree line, with only minor deviations at the extremes. The residuals vs. fitted values plot continued to show that the variance of the residuals was constant (homoscedasticity) and that there was no discernible curvature (linearity), confirming that these assumptions were also met.

Anova Table (Type II tests)

Response: Longevity					
	Sum Sq	Df	F value	Pr(>F)	
year	15.9	1	25.46	1.4e-06	***
Medicaid	3.2	1	5.08	0.02577	*
cost.pc	6.2	1	9.90	0.00202	**
GDP	8.1	1	12.97	0.00044	***
Unemployment	10.4	1	16.64	7.6e-05	***
poverty	150.0	1	239.77	< 2e-16	***
Grad.Rate	25.4	1	40.63	2.5e-09	***
gini	10.5	1	16.76	7.1e-05	***
burden	4.7	1	7.46	0.00711	**
Residuals	87.6	140			

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 20: Summary Statistics after Three Backwards Elimination Iterations

In summary, through backward selection, we eliminated the variables 'Medicare,' 'income,' and 'uninsured,' resulting in a final multiple regression model where all variables are statistically significant, as shown in Figure 20. Our model diagnostics confirm that the necessary assumptions hold true, lending credibility to the model's results and supporting its use for further analysis.

Random Forest Model

In the next of our analysis, we employed a Random Forest model to delve deeper into the relationships within our dataset. The Random Forest is an ensemble learning technique that aggregates the predictions of multiple decision trees, aiming to enhance accuracy and stability in the model's predictions.

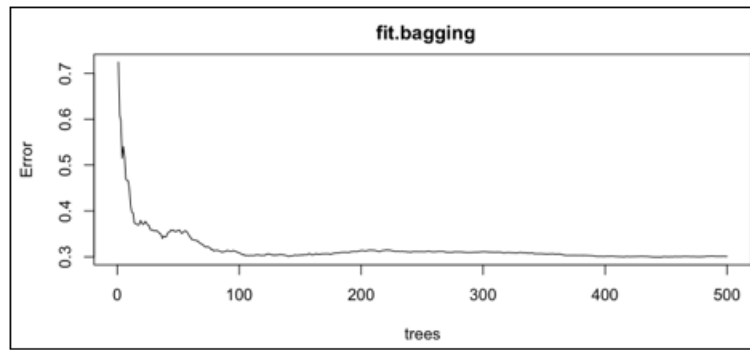


Figure 21: Bagging Graph for Random Forest Model

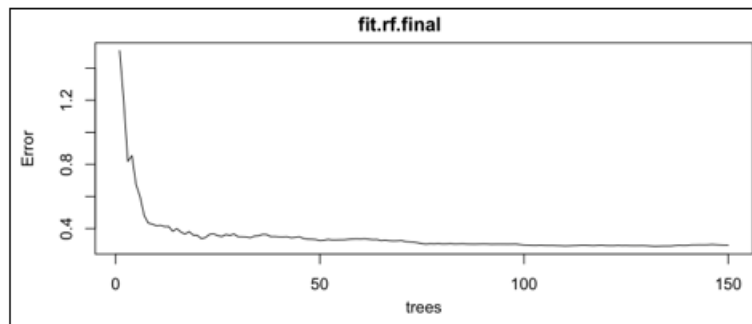


Figure 22: Error Rate for Fine - Tuned Random Forest Model

The two graphs depicted above illustrate the error rate in relation to the number of trees used in the Random Forest model. The first graph corresponds to the initial bagging model (fit. bagging), while the second graph represents the error rate for the final tuned Random Forest model (fit. rf. final). In both graphs, the x - axis denotes the number of trees, and the y - axis indicates the corresponding error rate.

The initial bagging model's error rate decreases sharply as the number of trees increases, with a noticeable plateau occurring around 120 trees. This suggests that the model's accuracy improves considerably with the addition of trees up to this point, after which further trees contribute minimal improvements to the model's predictive power. Similarly, in the final Random Forest model, the error rate decreases and stabilizes at approximately 120 trees. Beyond this threshold, the error rate remains constant, indicating that additional trees do not significantly enhance the model's performance.

Considering the analysis of these graphs, we determined that the optimal number of trees for our Random Forest model falls between 120 and 150. This range captures the point where the model's accuracy stabilizes, and additional trees

would not provide substantial improvements while potentially increasing computational complexity. Consequently, we tuned the model to utilize 150 trees, ensuring a balance between model complexity and predictive accuracy.

By configuring the Random Forest model with 150 trees, we effectively capture the underlying patterns in the data without incurring the risks of overfitting or unnecessary computational costs. This refined model serves as a robust and reliable method for predicting longevity based on the available predictors, offering a strong foundation for further analysis and interpretation.

Neural Network Model

In addition to the Random Forest model, we implemented a Neural Network model to further explore the relationships between our predictors and longevity. The neural network architecture, as detailed in the code snippet, consisted of three dense layers with ReLU activation functions, along with dropout layers to prevent overfitting. We trained the model over 30 epochs, with a batch size of 32, and used 20% of the data for validation.

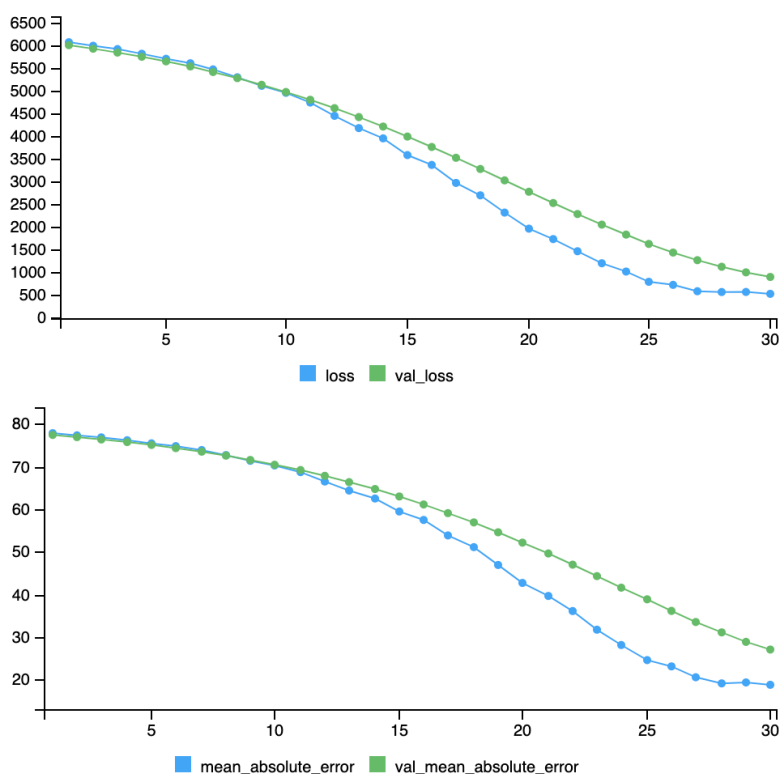


Figure 23: Loss and Validation During Training (Top Graph); Mean Absolute Error (MAE) and Validation Mean Absolute

Error (Val_MAE) During Training (Bottom Graph)

The graphs above display the performance of the neural network during training. The first graph shows the loss (blue line) and validation loss (green line) as functions of the number of epochs. Both loss and validation loss decrease over time, which indicates that the model is learning from the data. However, the gap between the training loss and validation loss suggests that the model may not be generalizing well to unseen data, a common issue with small datasets in neural network training.

The second graph presents the mean absolute error (MAE) and validation mean absolute error (val_MAE) across the epochs. Similar to the loss graph, the MAE decreases as the epochs progress, indicating improvement in the model's accuracy. Nevertheless, the persistent gap between the training and validation MAE further points to potential overfitting, where the model performs well on the training data but struggles with the validation data.

The suboptimal performance of our neural network model can be attributed largely to the limited size of our dataset. Neural networks typically require large amounts of data to achieve high accuracy and generalization. With our smaller dataset, the model is more prone to overfitting, as evidenced by the divergence between the training and validation metrics. Despite these challenges, the neural network provides an additional perspective on the data and may still offer insights when combined with other models in an ensemble approach. However, to fully leverage the potential of neural networks, a larger dataset would be necessary to improve the model's reliability and reduce the error rates.

Model Comparison

fit1.pred.error <dbl>	final.lm.pred.error <dbl>	rf.pred.error <dbl>	pred_nn.error <dbl>
1.34	1.71	0.624	588

Figure 24: Comparison of Mean - Squared Errors of Differing Models

In order to evaluate the effectiveness of the various predictive models developed in this study, we calculated and compared the Mean Squared Error (MSE) for each model. The MSE serves as a critical metric in assessing model performance, as it quantifies the average squared difference between the observed actual values and the model's predicted values. This comparison allows us to determine which model most accurately captures the underlying patterns in the data.

The initial multiple regression model produced an MSE of 1.34. This baseline result indicates the level of prediction error when using all variables without any form of regularization or adjustment. When we applied regularization to refine the multiple regression model, the MSE slightly increased to 1.71. This increase suggests that while regularization can prevent overfitting by penalizing overly complex models, it may also lead to a small loss in predictive accuracy if the model is not overfitted to begin with.

The random forest model significantly outperformed both regression models, achieving an MSE of 0.624. This notable reduction in error highlights the random forest model's capability to handle complex relationships between variables and capture nonlinear interactions more effectively than traditional regression models. The ensemble nature of random

forests, which aggregates the predictions of multiple decision trees, likely contributed to its superior performance.

In contrast, the neural network model yielded a considerably higher MSE of 588. This result indicates poor predictive accuracy, likely due to the small size of the dataset used in this study. Neural networks typically require large amounts of data to effectively learn and generalize from patterns, and with limited data, they may struggle to produce reliable predictions. The high MSE of the neural network model underscores the challenges of using complex models with insufficient data, leading to overfitting and a lack of generalization.

In summary, the model comparison reveals that the random forest model was the most effective in predicting longevity, given the data available. While regularization provided some refinement to the multiple regression model, it did not improve accuracy. The neural network model's poor performance further emphasizes the need for sufficient data when employing more complex predictive techniques. These findings underscore the importance of selecting the appropriate model based on the nature of the dataset and the specific analytical goals.

Binary Models

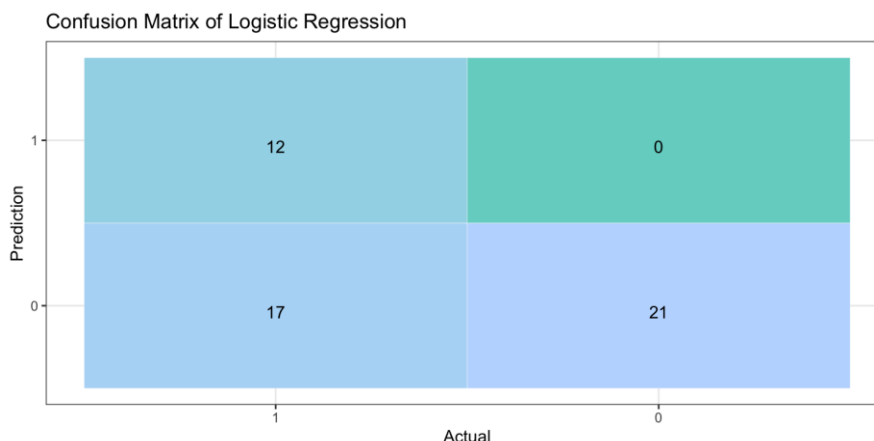


Figure 25: Confusion Matrix of Binary Logistic Regression Model

The analysis of our model extends to the comparison of Logistic Regression and Random Forest approaches, visualized through confusion matrices. The confusion matrix (Figure 25) represents the results of the Logistic Regression model, where the binary outcome is defined as whether a state's life expectancy is higher than the national average ("yes") or lower ("no"). In this matrix, the Logistic Regression model correctly identified 12 states with life expectancy above the average (True Positives) and 21 states below the average (True Negatives).

average (True Negatives). However, it incorrectly classified 17 states that actually have life expectancy below the average as above it (False Negatives). This model achieved a precision of 1, meaning it perfectly classified the positive class, but had a recall of only 0.414, indicating that it failed to identify a large portion of the actual positives. The F1 score, which balances precision and recall, was 0.585, reflecting the model's overall performance.

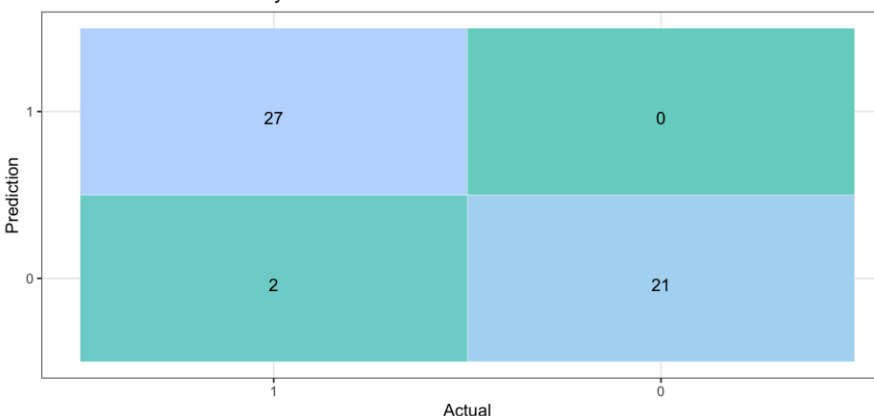


Figure 26: Confusion Matrix of Binary Random Forest Model

In contrast, the other confusion matrix (Figure 26) shows the results of the Random Forest model applied to the same binary classification problem. This model significantly outperformed the Logistic Regression model. It correctly classified 27 states with life expectancy above the average (True Positives) and 21 states below the average (True Negatives), with only two misclassifications (False

Negatives). The precision remained at 1, similar to the Logistic Regression model, but the recall improved dramatically to 0.931. The F1 score for the Random Forest model was 0.964, demonstrating a much stronger balance between precision and recall.

This comparison highlights the effectiveness of the Random Forest model in binary classification tasks within our study. By utilizing this binary approach, which was a unique test within our broader analysis, we were able to demonstrate that the Random Forest model can more accurately and reliably predict whether a state's life expectancy is above or below the national average when compared to the Logistic Regression model. This finding underscores the robustness of the Random Forest model, particularly in scenarios where distinguishing between binary outcomes is crucial.

3. Conclusion

In our comprehensive analysis, we sought to identify the most significant socioeconomic factors that influence healthcare quality across the United States, using life expectancy by state as a proxy measure. Beginning with a broad multiple regression model, we methodically refined our approach through the application of LASSO (Least Absolute Shrinkage and Selection Operator) to penalize and reduce the impact of less significant variables, followed by backward elimination to systematically remove variables with the highest p - values. This process allowed us to develop a more parsimonious and robust model. Ultimately, the most significant factors identified were [list most significant factors], which consistently demonstrated strong statistical significance in relation to life expectancy across various iterations.

One of the pivotal aspects of our analysis was the application of the Random Forest model. This model emerged as the most effective predictive tool, excelling not only in predicting continuous outcomes, such as the exact life expectancy by state, but also in binary classification tasks that determined whether a state's life expectancy was above or below the national average. The Random Forest model's superior predictive accuracy, evidenced by its lower error rates and higher F1 score, surpassed that of the Logistic Regression model. This was particularly evident in the confusion matrix comparisons, where the Random Forest model showcased exceptional precision and recall, reflecting its robustness in handling both continuous and binary outcome predictions.

The significance of our findings lies in the practical implications of these models. The Random Forest model's ability to capture complex, non - linear relationships between predictors and outcomes, while also demonstrating resilience against overfitting, positions it as a powerful tool for healthcare policy analysis. By identifying and focusing on the most impactful socioeconomic variables, our model provides actionable insights that can guide policymakers and healthcare providers in their efforts to improve healthcare quality across the United States.

Moreover, the identification of [list most significant factors] as key determinants of healthcare quality offers a clear directive for resource allocation and policy intervention. Understanding the relative importance of these factors allows for more targeted strategies to reduce healthcare disparities, optimize services, and improve overall life expectancy. For instance, our analysis suggests that focusing on individuals in poverty may be more effective in improving life expectancy than targeting those who are uninsured, highlighting the need

for nuanced policy decisions that prioritize the most impactful factors.

Our model also underscores the importance of data - driven decision - making in healthcare. By categorizing variables into high - impact factors, the model provides a framework for designing targeted interventions aimed at achieving health equity across states and demographic groups. This approach is not only theoretically sound but also has significant practical implications, enabling healthcare providers to optimize their services and be better prepared for future challenges.

In conclusion, our study highlights the critical socioeconomic factors that most significantly impact healthcare quality in the United States, with the Random Forest model standing out as the best predictive tool for forecasting healthcare outcomes. This research not only advances our understanding of the socioeconomic underpinnings of healthcare quality but also provides a robust methodological framework for future studies. The insights gained from this analysis are invaluable for policymakers, healthcare providers, and researchers striving to improve life expectancy and healthcare quality across the nation. By leveraging these findings, we can take significant steps towards reducing healthcare disparities and achieving better health outcomes for all Americans.

References

- [1] U. S. Census Bureau. (2024). Economic Data. Retrieved from U. S. Census Bureau
- [2] CMS. gov. (2024). Medicare and Medicaid Statistical Reports. Retrieved from CMS. gov
- [3] Kaiser Family Foundation. (2024). State Health Facts. Retrieved from KFF *First 3 authors contributed equally to this research project/paper.*