

# An Analysis of Factors Affecting County-Level Suicide Rates

Jenna Gorham

## Introduction and Methods

Suicide is a significant public health concern. In the United States, suicide is the 12th leading cause of death with an age-adjusted suicide rate of 13.48 per 100,000 individuals (<https://afsp.org/suicide-statistics/>). However, suicide is also known to be preventable. Suicide rates vary across U.S. counties, and so using statistical analysis and the tools of linear regression, county-level suicide rates can be modeled as a function of various socioeconomic and demographic characteristics of those counties. This analysis can help generate insight into those factors that are most predictive of high suicide rates with the goal of improving policy for suicide prevention.

The data from this project is from County Health Rankings and Roadmaps (CHR&R), a program of the University of Wisconsin Population Health Institute (<https://www.countyhealthrankings.org/>). CHR&R collects data from a variety of original sources to create a dataset summarizing information on healthcare outcomes, socioeconomic factors, and demographic factors in every U.S. county. For this project, we utilize the 2023 dataset, focusing on the variables of age-adjusted suicide rates and relevant socioeconomic and demographic factors that may be of interest to researchers and policymakers. In total, this gives us 30 variables for analysis, one response variable and 29 predictors.

## Analysis

We begin with data visualization. First, we wish to get a sense of the distribution of suicide rates across counties. The two visualizations that are most helpful for us in this regard are a histogram and a boxplot (see Appendix A).

We see a heavily right-skewed distribution with several outlier counties in which suicide rates were exceptionally high. The presence of these unusual points may negatively impact linear models fit to the data.

Next, we turn towards exploring the effect of individual predictors on the suicide rates. Unfortunately, with 29 predictors, visualizing the effect of each predictor is not practical. We achieve this exploration through summary statistics and data analysis. For each predictor, we first scale the predictor for ease of comparison. Next, we compute its correlation with suicide rate. Then we fit a simple linear regression model using the scaled predictor to model suicide rate, and we record the slope coefficient and  $p$ -value associated with the predictor. We also record the  $R^2$  statistic from the linear model. These statistics allow us to determine the strength and nature of the relationship between the predictors. These values are summarized in a table below, sorted by ascending  $p$ -value, such that those predictors that result in statistically significant slope coefficients are listed above those that do not. (See Appendix B for the table.)

This analysis immediately gives us some idea of which predictors might be important in modeling suicide rates. The number of firearm fatalities in a county is highly predictive of suicide rates. Some other significant predictors include how rural a country is, measures of poverty, and demographic features. Some insignificant predictors include income inequality and the number of mental health providers in a county.

## Linear Models

Having seen that certain predictors are likely not useful in predicting suicide rates, fitting a full multiple linear regression model is likely a naive approach. Instead, we use a criterion model selection algorithm based on the AIC (Akaike information criterion) criterion, iterating through all possible linear models that can be fit from the 29 predictors and selecting the model that results in the lowest AIC. (See Appendix C for implementation and linear model summary).

The resultant linear model is given below:

$$\begin{aligned} \widehat{suicides} = & 23.79 - 1.048(food\_environment\_index) + 1.427(access\_to\_exercise) \\ & + 19.23(unsured) + 272.0(mental\_health\_providers) \\ & + 19.98(high\_school\_completion\_rate) - 0.3933(income\_inequality) \\ & - 0.1082(social\_associations) - 0.2451(air\_pollution) \\ & + 16.63(severe\_housing\_problems) + (2.229 \times 10^{-4})(school\_funding\_adequacy) \\ & - 0.04707(nonwhite\_white\_residential\_segregation) \\ & - 7.274(broadband\_access) + 52.98(below\_18) + 39.81(65\_and\_over) \\ & + 8.723(white\_pop) - 75.47(female\_pop) - 18.25(children\_poverty) \\ & - (4.569 \times 10^{-5})(median\_income) + 0.5806(firearm\_fatalities). \end{aligned}$$

Note that this model shows some unexpected results; for example, it predicts that as the number of

mental health providers in a county increases, the suicide rate increases. This is interesting, so we investigate it further by visualizing the relationship between mental health providers and suicides in a scatterplot (see Appendix D). There does not appear to be any real relationship between the two variables, as our initial data exploration suggested, but there are some extreme outliers in the data.

Now we turn to investigating the quality of fit. To do so, we use two tools: a QQ-plot and a plot of the residual versus the fitted values of the model. (See Appendix E for both plots.) We see several immediate issues. The errors do not appear to be normally distributed, with the QQ plot indicating a heavy-tailed distribution. The plot of fitted and residual values indicates non-constant variance and the presence of multiple outliers.

To attempt to fit a better model, we will apply some transformation methods. Because our issues with the original model include both non-normality and non-constant variance, we should likely need to transform both the response (suicide rate) and some predictor variables. Since we observed a “megaphone”-like shape in the residual vs. fit plot, we will initially chose the square root transformation for the response. To determine the transformations, if any, we should perform on the predictors, we individually visualize the square root of suicides as a function of each predictors, and if the relationship appears nonlinear, we choose the transformation that results in the best linear relationship between predictor and response. (For the sake of space, this process has been omitted from the appendixes.) We determine that we should take the following transformations: the log of uninsured, primary care physicians, mental health providers, unemployment, social associations, population, and median income. More precisely, for each of these predictors, we perform the transformation  $\log(x + 1)$  to avoid an error when taking the log of zero.

Given these transformations, we fit a full linear model including all terms and then proceed to use the AIC criterion selection algorithm as before to determine the best subset of the terms to include in the model (see Appendix F). The resultant model is given below:

$$\begin{aligned}
\sqrt{\widehat{suicides}} = & 9.297 - 0.1129(food\_environment\_index) + 0.1388(access\_to\_exercise) \\
& + 2.849 \log(uninsured + 1) + 31.91 \log(mental\_health\_providers + 1) \\
& + 2.225(high\_school\_completion\_rate) - 0.06188(income\_inequality) \\
& - 0.1314 \log(social\_associations + 1) - 0.01710(air\_pollution) \\
& + 0.03656(drinking\_water\_violations) + 1.530(severe\_housing\_problems) \\
& + (1.925 \times 10^{-5})(school\_funding\_adequacy) \\
& - (3.575 \times 10^{-3})(nonwhite\_white\_residential\_segregation) \\
& - 0.4166(broadband\_access) - 0.04577 \log(population + 1) \\
& + 5.020(below\_18) + 3.771(65\_and\_over) + 1.061(white\_pop) \\
& - 7.538(female\_pop) - 0.02215(children\_poverty) \\
& - 0.4170 \log(median\_income + 1) + 0.06066(firearm\_fatalities).
\end{aligned}$$

Similar to our first model, this model includes unintuitive results that seem to differ from what we see in scatterplots of the data. Perhaps this phenomenon is a result of influential points and outliers. Perhaps if outliers were removed, there would be fewer terms in the model as well for this reason. This might improve our ability to make inferences about factors that contribute to higher suicide rates.

As before, we investigate the quality of fit with a QQ-plot and a plot of the residuals versus the fitted values (see Appendix G). The QQ-plot looks much closer to a normal distribution than the QQ-plot for the original linear model. The residual versus fitted plot indicates that there is still some concern for non-constant variance. One additional metric we can use to compare this model to the original model is the adjusted  $R^2$  values associated with each model. The first linear model has an adjusted  $R^2$  of 0.6397, and the transformed model has an adjusted  $R^2$  of 0.6492. These values are comparable, so we did not lose explanatory power by transforming our variables. In fact, adjusted  $R^2$  is higher for the transformed model, perhaps indicating that our transformations allowed for a better linear fit between the predictors and response.

This model appears to be an improvement, but it still has some issues. Perhaps a different choice of transformation would result in a better fitting model, or perhaps removing the most extreme of unusual point would help. Alternatively, it is possible that we would obtain a better model by using a different model selection algorithm. Further exploration and trial-and-error would be necessary here to determine how to obtain the best possible model.

## Discussion and Conclusion

Initial data exploration showed that individually, some variables are better at predicting suicide rates than others. We began by constructing a naive linear model and selected a subset of predictors that minimized

AIC. This model did not perform well with poor diagnostics. Transforming the response variable and some of the predictors resulted in a linear model with the same explanatory power and better diagnostics. Further improvements could be made by trying different transformation methods, removing outliers, trying a more flexible model like quadratic regression, or implementing weighted least squares to address the issue of non-constant variance. These models could be further evaluated by using cross-validation to split the data into training and testing sets to determine how the models perform in terms of making predictions.

Certain variables were consistently seen to be strongly predictive of high suicide rates. For example, both initial data exploration and both linear models showed firearm fatalities to be positively associated with suicides. This may indicate that third variable is a cause of both firearm fatalities and high suicide rates - namely, easy access to firearms. So perhaps an area wishing to prevent deaths by suicides might be interested in restricting access to firearms or other lethal weapons. Certain demographic factors also appear significant in predicting suicides; for example, a larger elderly population is associated with higher suicide rates and a larger female population is associated with lower suicide rates. This is likely reflective of the demographics of individuals more likely to die by suicide.

Not all predictors were so clear cut. Notably, we discussed the impact of the number of mental health providers on suicide rates. Initial data exploration indicated no predictive relationship, and a scatterplot of the two variables appears to corroborate this, but in both linear models, an increase in mental health providers is predicted to increase the number of suicides in a county. This also runs counter to both intuition and prior knowledge. Further study is required to determine the true nature of this relationship in the data. It is possible that this term appears in the models due to the impact of extreme observations when no such relationship truly exists. Alternatively, it is possible that the term appears as a result of an interaction between mental health providers and some other variable.

## Appendix A: Data Visualization and R Code

```
p1 <- suicides %>%
  ggplot(aes(x = suicides)) +
  geom_histogram() +
  theme_bw()

p2 <- suicides %>%
  ggplot(aes(y = suicides)) +
  geom_boxplot() +
  theme_bw()

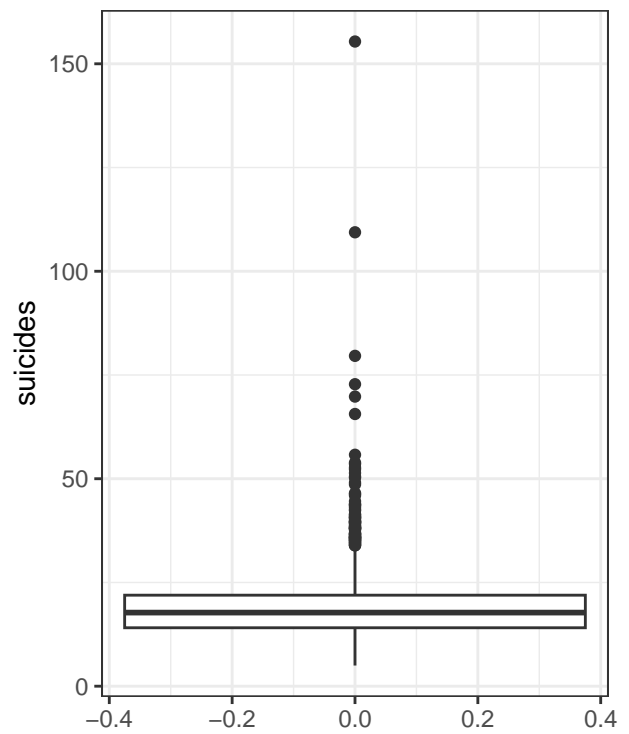
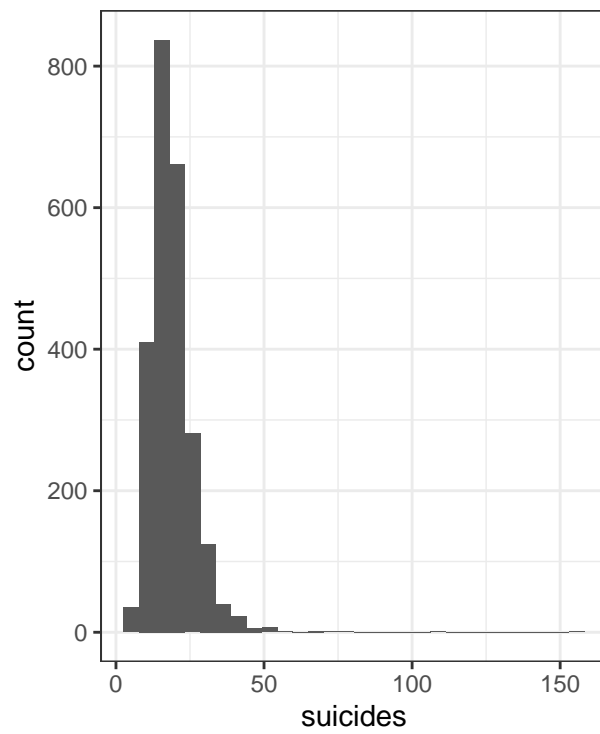
p3 <- plot_grid(p1, p2)

title_theme <- calc_element("plot.title", theme_bw())

title_obj <- ggdraw() +
  draw_label(
    "Distribution of County-Level Suicide Rates",
    fontfamily = title_theme$family,
    fontface = title_theme$face,
    size = title_theme$size,
    hjust = 0,
    x = 0.05
  )

plot_grid(title_obj, p3, nrow = 2, ncol = 1, rel_heights = c(0.15, 1))
```

## Distribution of County-Level Suicide Rates



## Appendix B: Data Exploration Table and R Code

```
# get list of the names of all predictors in the dataset
predictors <- colnames(suicides)
predictors <- predictors[!(predictors == 'suicides')]

r2 <- correlation <- coefficient <- pval <- numeric(length(predictors))

for (i in 1:length(predictors)) {
  varname <- predictors[i] # get the name of the predictor we are working with
  varcol <- as.vector(unname(suicides[varname]))[[1]] # get column for that var

  # next, we scale varcol
  mean_varcol <- mean(varcol, na.rm = TRUE)
  sd_varcol <- sd(varcol, na.rm = TRUE)
  varcol_scaled = ifelse(is.na(varcol), varcol, (varcol - mean_varcol) / sd_varcol)
  # where varcol is NA, we keep it as NA
  # where varcol has a numeric value, we scale it (convert to z-score)

  # next, we create a dataframe with our relevant variables
  df <- data.frame(x = as.vector(varcol_scaled), y = suicides$suicides)
  # now we can fit our simple linear model
  linmod <- lm(y ~ ., data = df)
  # we get the data we wanted from the summary
  pval[i] <- summary(linmod)$coefficients[2, 4]
  coefficient[i] <- summary(linmod)$coefficients[2, 1]
  r2[i] <- summary(linmod)$r.squared

  # now, we compute the correlation
  correlation[i] <- cor(na.omit(df)$x, na.omit(df)$y)
}

# summarize our results
results <- data.frame(Predictor = predictors,
                      Correlation = correlation,
                      Coefficient = coefficient,
                      pvalue = pval,
                      Rsquared = r2) %>%

  arrange(pvalue)
kable(results)
```

Predictor	Correlation	Coefficient	pvalue	Rsquared
firearm_fatalities	0.6168405	4.5421546	0.0000000	0.3804922
percent_rural	0.3382691	2.7996376	0.0000000	0.1144260
median_income	-0.2834087	-2.1474231	0.0000000	0.0803205
food_environment_index	-0.2802648	-2.4723336	0.0000000	0.0785484
food_insecurity	0.2771645	2.2860561	0.0000000	0.0768202
broadband_access	-0.2493997	-2.1233536	0.0000000	0.0622002
air_pollution	-0.2352115	-1.7541795	0.0000000	0.0553244
65_and_over	0.2296428	1.9138649	0.0000000	0.0527358
population	-0.2261210	-1.5876842	0.0000000	0.0511307
access_to_exercise	-0.2260760	-1.8203417	0.0000000	0.0511104

Predictor	Correlation	Coefficient	pvalue	Rsquared
female_pop	-0.1909886	-1.8287985	0.0000000	0.0364767
uninsured	0.1853202	1.5240071	0.0000000	0.0343436
children_poverty	0.1831927	1.5442413	0.0000000	0.0335596
homeownership_rate	0.1467449	1.1772326	0.0000000	0.0215341
nonwhite_white_residential_segregation	-0.1401664	-1.0623938	0.0000000	0.0196466
primary_care_physicians	-0.1299086	-1.0440706	0.0000000	0.0168762
school_segregation	-0.1277014	-0.9425599	0.0000000	0.0163077
white_pop	0.1134989	0.9371724	0.0000000	0.0128820
gender_pay_gap	-0.0775492	-0.7331586	0.0001287	0.0060139
below_18	0.0747804	0.6194713	0.0002226	0.0055921
childcare_cost_burden	0.0733452	0.5699433	0.0002937	0.0053795
social_associations	-0.0521094	-0.5635622	0.0101480	0.0027154
severe_housing_problems	0.0433459	0.3525804	0.0325197	0.0018789
unemployment	-0.0404634	-0.3632655	0.0459699	0.0016373
high_school_completion_rate	-0.0319246	-0.2805223	0.1154207	0.0010192
drinking_water_violations	0.0267546	0.2062619	0.1903876	0.0007158
income_inequality	0.0240901	0.2131720	0.2353881	0.0005803
mental_health_providers	0.0186423	0.1500672	0.3598054	0.0003475
school_funding_adequacy	-0.0102574	-0.0776997	0.6157064	0.0001052

## Appendix C: Linear Model Selection

```
library(bestglm)
```

```
## Loading required package: leaps
```

```
# the data needs to be formatted very specifically here
```

```
suicides2 <- suicides
```

```
names(suicides2)[30] <- 'y'
```

```
suicides2 <- as.data.frame(suicides2)
```

```
bestglm(na.omit(suicides2), IC = 'AIC')$Subsets
```

```
##      (Intercept) food_environment_index access_to_exercise uninsured
## 0             TRUE                FALSE                FALSE      FALSE
## 1             TRUE                FALSE                FALSE      FALSE
## 2             TRUE                FALSE                FALSE      FALSE
## 3             TRUE                FALSE                FALSE      FALSE
## 4             TRUE                FALSE                FALSE       TRUE
## 5             TRUE                FALSE                FALSE       TRUE
## 6             TRUE                FALSE                FALSE      FALSE
## 7             TRUE                 TRUE                FALSE      FALSE
## 8             TRUE                FALSE                FALSE       TRUE
## 9             TRUE                FALSE                FALSE       TRUE
## 10            TRUE                TRUE                FALSE       TRUE
## 11            TRUE                TRUE                FALSE       TRUE
## 12            TRUE                TRUE                FALSE       TRUE
## 13            TRUE                TRUE                FALSE       TRUE
## 14            TRUE                TRUE                FALSE       TRUE
## 15            TRUE                TRUE                FALSE       TRUE
## 16            TRUE                TRUE                FALSE       TRUE
## 17            TRUE                TRUE                FALSE       TRUE
## 18            TRUE                TRUE                FALSE       TRUE
## 19*           TRUE                TRUE                 TRUE       TRUE
## 20            TRUE                TRUE                 TRUE       TRUE
## 21            TRUE                TRUE                 TRUE       TRUE
## 22            TRUE                TRUE                 TRUE       TRUE
## 23            TRUE                TRUE                 TRUE       TRUE
## 24            TRUE                TRUE                 TRUE       TRUE
## 25            TRUE                TRUE                 TRUE       TRUE
## 26            TRUE                TRUE                 TRUE       TRUE
## 27            TRUE                TRUE                 TRUE       TRUE
## 28            TRUE                TRUE                 TRUE       TRUE
## 29            TRUE                TRUE                 TRUE       TRUE
##      primary_care_physicians mental_health_providers high_school_completion_rate
## 0                FALSE                FALSE                FALSE
## 1                FALSE                FALSE                FALSE
## 2                FALSE                FALSE                FALSE
## 3                FALSE                FALSE                FALSE
## 4                FALSE                FALSE                FALSE
## 5                FALSE                FALSE                FALSE
## 6                FALSE                FALSE                TRUE
## 7                FALSE                FALSE                FALSE
```

## 8	FALSE	FALSE	TRUE	
## 9	FALSE	TRUE	TRUE	
## 10	FALSE	FALSE	TRUE	
## 11	FALSE	TRUE	TRUE	
## 12	FALSE	TRUE	TRUE	
## 13	FALSE	TRUE	TRUE	
## 14	FALSE	TRUE	TRUE	
## 15	FALSE	TRUE	TRUE	
## 16	FALSE	TRUE	TRUE	
## 17	FALSE	TRUE	TRUE	
## 18	FALSE	TRUE	TRUE	
## 19*	FALSE	TRUE	TRUE	
## 20	FALSE	TRUE	TRUE	
## 21	FALSE	TRUE	TRUE	
## 22	FALSE	TRUE	TRUE	
## 23	FALSE	TRUE	TRUE	
## 24	FALSE	TRUE	TRUE	
## 25	FALSE	TRUE	TRUE	
## 26	FALSE	TRUE	TRUE	
## 27	FALSE	TRUE	TRUE	
## 28	FALSE	TRUE	TRUE	
## 29	TRUE	TRUE	TRUE	
##	unemployment	income_inequality	social_associations	air_pollution
## 0	FALSE	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE	FALSE
## 4	FALSE	FALSE	FALSE	FALSE
## 5	FALSE	FALSE	FALSE	FALSE
## 6	FALSE	FALSE	FALSE	FALSE
## 7	FALSE	FALSE	FALSE	FALSE
## 8	FALSE	FALSE	FALSE	FALSE
## 9	FALSE	FALSE	FALSE	FALSE
## 10	FALSE	FALSE	FALSE	FALSE
## 11	FALSE	FALSE	FALSE	FALSE
## 12	FALSE	FALSE	FALSE	FALSE
## 13	FALSE	FALSE	FALSE	FALSE
## 14	FALSE	FALSE	FALSE	TRUE
## 15	FALSE	FALSE	TRUE	TRUE
## 16	FALSE	FALSE	TRUE	TRUE
## 17	FALSE	FALSE	TRUE	TRUE
## 18	FALSE	TRUE	TRUE	TRUE
## 19*	FALSE	TRUE	TRUE	TRUE
## 20	FALSE	TRUE	TRUE	TRUE
## 21	FALSE	TRUE	TRUE	TRUE
## 22	TRUE	TRUE	TRUE	TRUE
## 23	TRUE	TRUE	TRUE	TRUE
## 24	TRUE	TRUE	TRUE	TRUE
## 25	TRUE	TRUE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE	TRUE
##	drinking_water_violations	severe_housing_problems	food_insecurity	

## 0	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE
## 4	FALSE	FALSE	FALSE
## 5	FALSE	FALSE	FALSE
## 6	FALSE	FALSE	FALSE
## 7	FALSE	FALSE	FALSE
## 8	FALSE	FALSE	FALSE
## 9	FALSE	FALSE	FALSE
## 10	FALSE	FALSE	FALSE
## 11	FALSE	FALSE	FALSE
## 12	FALSE	FALSE	FALSE
## 13	FALSE	TRUE	FALSE
## 14	FALSE	TRUE	FALSE
## 15	FALSE	FALSE	FALSE
## 16	FALSE	TRUE	FALSE
## 17	FALSE	TRUE	FALSE
## 18	FALSE	TRUE	FALSE
## 19*	FALSE	TRUE	FALSE
## 20	FALSE	TRUE	TRUE
## 21	FALSE	TRUE	TRUE
## 22	FALSE	TRUE	TRUE
## 23	FALSE	TRUE	TRUE
## 24	TRUE	TRUE	TRUE
## 25	TRUE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE
##	school_segregation	school_funding_adequacy	gender_pay_gap
## 0	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE
## 4	FALSE	TRUE	FALSE
## 5	FALSE	TRUE	FALSE
## 6	FALSE	FALSE	FALSE
## 7	FALSE	FALSE	FALSE
## 8	FALSE	TRUE	FALSE
## 9	FALSE	TRUE	FALSE
## 10	FALSE	TRUE	FALSE
## 11	FALSE	TRUE	FALSE
## 12	FALSE	TRUE	FALSE
## 13	FALSE	TRUE	FALSE
## 14	FALSE	TRUE	FALSE
## 15	FALSE	TRUE	FALSE
## 16	FALSE	TRUE	FALSE
## 17	FALSE	TRUE	FALSE
## 18	FALSE	TRUE	FALSE
## 19*	FALSE	TRUE	FALSE
## 20	FALSE	TRUE	FALSE
## 21	FALSE	TRUE	FALSE
## 22	FALSE	TRUE	FALSE

## 23	FALSE	TRUE	FALSE		
## 24	FALSE	TRUE	FALSE		
## 25	FALSE	TRUE	FALSE		
## 26	FALSE	TRUE	FALSE		
## 27	TRUE	TRUE	FALSE		
## 28	TRUE	TRUE	TRUE		
## 29	TRUE	TRUE	TRUE		
##	nonwhite_white_residential_segregation	childcare_cost_burden			
## 0		FALSE	FALSE		
## 1		FALSE	FALSE		
## 2		FALSE	FALSE		
## 3		FALSE	FALSE		
## 4		FALSE	FALSE		
## 5		FALSE	FALSE		
## 6		FALSE	FALSE		
## 7		FALSE	FALSE		
## 8		FALSE	FALSE		
## 9		FALSE	FALSE		
## 10		TRUE	FALSE		
## 11		TRUE	FALSE		
## 12		TRUE	FALSE		
## 13		TRUE	FALSE		
## 14		TRUE	FALSE		
## 15		TRUE	FALSE		
## 16		TRUE	FALSE		
## 17		TRUE	FALSE		
## 18		TRUE	FALSE		
## 19*		TRUE	FALSE		
## 20		TRUE	FALSE		
## 21		TRUE	FALSE		
## 22		TRUE	FALSE		
## 23		TRUE	FALSE		
## 24		TRUE	FALSE		
## 25		TRUE	TRUE		
## 26		TRUE	TRUE		
## 27		TRUE	TRUE		
## 28		TRUE	TRUE		
## 29		TRUE	TRUE		
##	homeownership_rate	broadband_access	population	below_18	65_and_over
## 0	FALSE	FALSE	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE	FALSE	FALSE
## 4	FALSE	FALSE	FALSE	FALSE	FALSE
## 5	FALSE	FALSE	FALSE	FALSE	FALSE
## 6	FALSE	FALSE	FALSE	TRUE	TRUE
## 7	FALSE	FALSE	FALSE	TRUE	TRUE
## 8	FALSE	FALSE	FALSE	TRUE	TRUE
## 9	FALSE	FALSE	FALSE	TRUE	TRUE
## 10	FALSE	FALSE	FALSE	TRUE	TRUE
## 11	FALSE	FALSE	FALSE	TRUE	TRUE
## 12	FALSE	FALSE	FALSE	TRUE	TRUE
## 13	FALSE	FALSE	FALSE	TRUE	TRUE
## 14	FALSE	FALSE	FALSE	TRUE	TRUE

## 15	FALSE	FALSE	FALSE	TRUE	TRUE
## 16	FALSE	FALSE	FALSE	TRUE	TRUE
## 17	FALSE	TRUE	FALSE	TRUE	TRUE
## 18	FALSE	TRUE	FALSE	TRUE	TRUE
## 19*	FALSE	TRUE	FALSE	TRUE	TRUE
## 20	FALSE	TRUE	FALSE	TRUE	TRUE
## 21	TRUE	TRUE	FALSE	TRUE	TRUE
## 22	TRUE	TRUE	FALSE	TRUE	TRUE
## 23	TRUE	TRUE	FALSE	TRUE	TRUE
## 24	TRUE	TRUE	FALSE	TRUE	TRUE
## 25	TRUE	TRUE	FALSE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE	TRUE	TRUE
##	white_pop	female_pop	percent_rural	children_poverty	median_income
## 0	FALSE	FALSE	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE	FALSE	FALSE
## 2	TRUE	FALSE	FALSE	FALSE	FALSE
## 3	TRUE	TRUE	FALSE	FALSE	FALSE
## 4	TRUE	FALSE	FALSE	FALSE	FALSE
## 5	TRUE	TRUE	FALSE	FALSE	FALSE
## 6	TRUE	TRUE	FALSE	FALSE	FALSE
## 7	TRUE	TRUE	FALSE	TRUE	FALSE
## 8	TRUE	TRUE	FALSE	FALSE	FALSE
## 9	TRUE	TRUE	FALSE	FALSE	FALSE
## 10	TRUE	TRUE	FALSE	FALSE	FALSE
## 11	TRUE	TRUE	FALSE	FALSE	FALSE
## 12	TRUE	TRUE	FALSE	TRUE	FALSE
## 13	TRUE	TRUE	FALSE	TRUE	FALSE
## 14	TRUE	TRUE	FALSE	TRUE	FALSE
## 15	TRUE	TRUE	FALSE	TRUE	TRUE
## 16	TRUE	TRUE	FALSE	TRUE	TRUE
## 17	TRUE	TRUE	FALSE	TRUE	TRUE
## 18	TRUE	TRUE	FALSE	TRUE	TRUE
## 19*	TRUE	TRUE	FALSE	TRUE	TRUE
## 20	TRUE	TRUE	FALSE	TRUE	TRUE
## 21	TRUE	TRUE	FALSE	TRUE	TRUE
## 22	TRUE	TRUE	FALSE	TRUE	TRUE
## 23	TRUE	TRUE	TRUE	TRUE	TRUE
## 24	TRUE	TRUE	TRUE	TRUE	TRUE
## 25	TRUE	TRUE	TRUE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE	TRUE	TRUE
##	firearm_fatalities	logLikelihood	AIC		
## 0	FALSE	-3868.943	7737.886		
## 1	TRUE	-3413.778	6829.556		
## 2	TRUE	-3216.265	6436.530		
## 3	TRUE	-3140.512	6287.024		
## 4	TRUE	-3102.293	6212.586		
## 5	TRUE	-3041.789	6093.578		
## 6	TRUE	-3011.468	6034.935		

```
## 7          TRUE      -2973.303 5960.606
## 8          TRUE      -2931.264 5878.528
## 9          TRUE      -2900.673 5819.346
## 10         TRUE      -2880.272 5780.544
## 11         TRUE      -2858.973 5739.946
## 12         TRUE      -2848.745 5721.490
## 13         TRUE      -2838.487 5702.974
## 14         TRUE      -2827.616 5683.232
## 15         TRUE      -2821.840 5673.680
## 16         TRUE      -2812.960 5657.919
## 17         TRUE      -2809.082 5652.164
## 18         TRUE      -2806.014 5648.029
## 19*        TRUE      -2803.081 5644.162
## 20         TRUE      -2802.569 5645.138
## 21         TRUE      -2802.116 5646.233
## 22         TRUE      -2801.457 5646.914
## 23         TRUE      -2800.971 5647.942
## 24         TRUE      -2800.583 5649.167
## 25         TRUE      -2800.323 5650.647
## 26         TRUE      -2800.120 5652.239
## 27         TRUE      -2799.986 5653.973
## 28         TRUE      -2799.920 5655.841
## 29         TRUE      -2799.918 5657.835
```

```
linmod <- lm(suicides ~ food_environment_index + access_to_exercise + uninsured
+ mental_health_providers + high_school_completion_rate + income_inequality
+ social_associations + air_pollution + severe_housing_problems
+ school_funding_adequacy + nonwhite_white_residential_segregation
+ broadband_access + below_18 + `65_and_over` + white_pop + female_pop
+ children_poverty + median_income + firearm_fatalities, data = suicides)
summary(linmod)
```

```
##
## Call:
## lm(formula = suicides ~ food_environment_index + access_to_exercise +
##     uninsured + mental_health_providers + high_school_completion_rate +
##     income_inequality + social_associations + air_pollution +
##     severe_housing_problems + school_funding_adequacy + nonwhite_white_residential_segregation +
##     broadband_access + below_18 + '65_and_over' + white_pop +
##     female_pop + children_poverty + median_income + firearm_fatalities,
##     data = suicides)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.3332  -2.1205  -0.1886   1.8936  23.8340
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.379e+01  4.085e+00   5.824 6.62e-09
## food_environment_index -1.048e+00  1.464e-01  -7.158 1.13e-12
## access_to_exercise    1.427e+00  5.463e-01   2.613 0.009046
## uninsured         1.923e+01  2.523e+00   7.622 3.75e-14
## mental_health_providers 2.720e+02  5.395e+01   5.041 5.02e-07
## high_school_completion_rate 1.998e+01  2.833e+00   7.052 2.38e-12
```

```

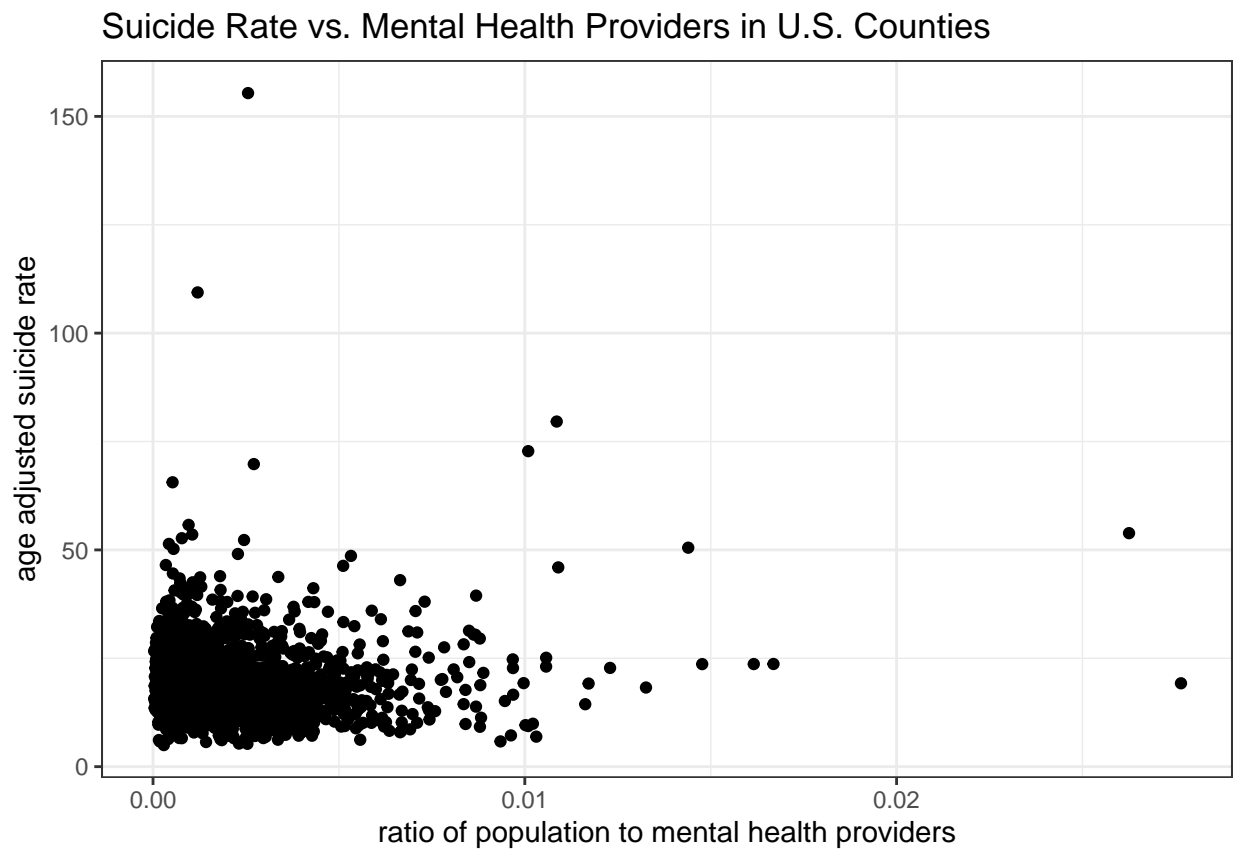
## income_inequality -3.933e-01 1.728e-01 -2.276 0.022949
## social_associations -1.082e-01 2.488e-02 -4.350 1.43e-05
## air_pollution -2.451e-01 5.609e-02 -4.369 1.31e-05
## severe_housing_problems 1.663e+01 3.552e+00 4.681 3.04e-06
## school_funding_adequacy 2.229e-04 2.936e-05 7.591 4.73e-14
## nonwhite_white_residential_segregation -4.707e-02 8.127e-03 -5.792 8.00e-09
## broadband_access -7.274e+00 1.993e+00 -3.650 0.000269
## below_18 5.398e+01 4.334e+00 12.457 < 2e-16
## '65_and_over' 3.981e+01 3.132e+00 12.710 < 2e-16
## white_pop 8.723e+00 7.970e-01 10.944 < 2e-16
## female_pop -7.547e+01 5.615e+00 -13.441 < 2e-16
## children_poverty -1.825e+01 2.787e+00 -6.547 7.35e-11
## median_income -4.569e-05 1.207e-05 -3.785 0.000158
## firearm_fatalities 5.806e-01 1.611e-02 36.030 < 2e-16
##
## (Intercept) ***
## food_environment_index ***
## access_to_exercise **
## uninsured ***
## mental_health_providers ***
## high_school_completion_rate ***
## income_inequality *
## social_associations ***
## air_pollution ***
## severe_housing_problems ***
## school_funding_adequacy ***
## nonwhite_white_residential_segregation ***
## broadband_access ***
## below_18 ***
## '65_and_over' ***
## white_pop ***
## female_pop ***
## children_poverty ***
## median_income ***
## firearm_fatalities ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.912 on 2102 degrees of freedom
## (1020 observations deleted due to missingness)
## Multiple R-squared: 0.643, Adjusted R-squared: 0.6397
## F-statistic: 199.2 on 19 and 2102 DF, p-value: < 2.2e-16

```

## Appendix D: Scatterplot of Suicides vs. Mental Health Providers

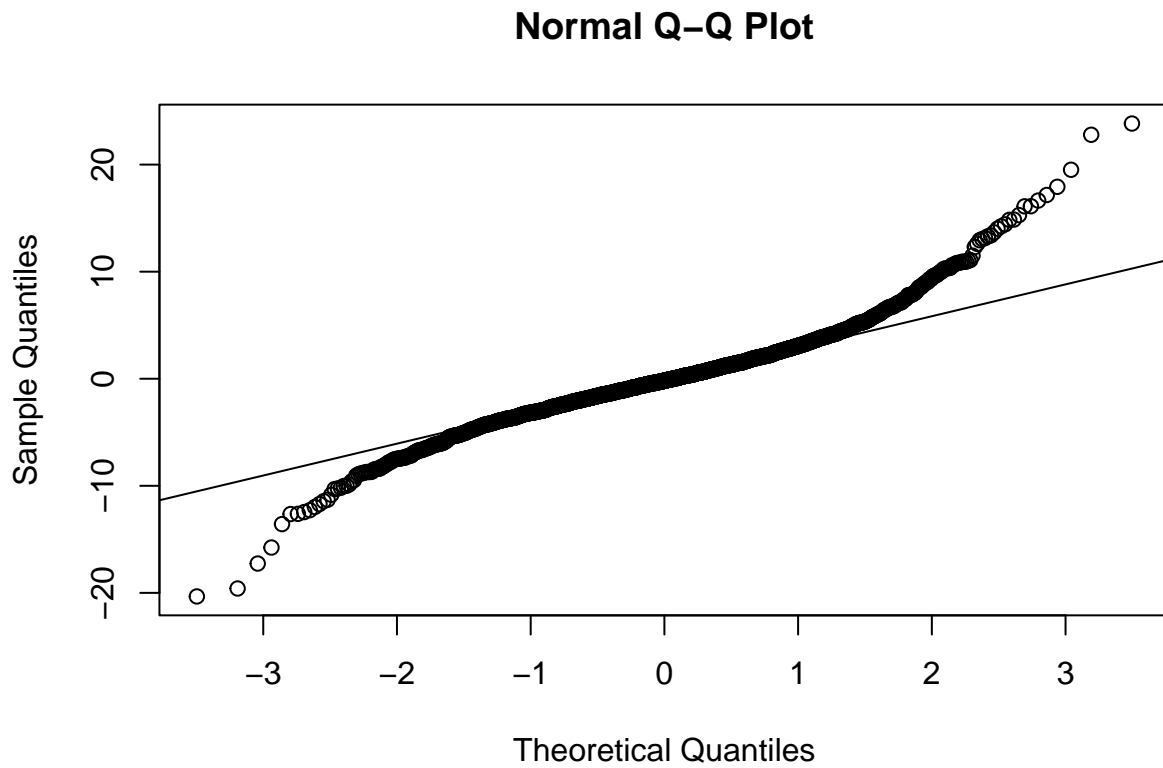
```
suicides %>%  
  ggplot(aes(x = mental_health_providers,  
             y = suicides)) +  
  geom_point() +  
  theme_bw() +  
  labs(title = 'Suicide Rate vs. Mental Health Providers in U.S. Counties',  
       x = 'ratio of population to mental health providers',  
       y = 'age adjusted suicide rate')
```

## Warning: Removed 727 rows containing missing values (‘geom\_point()’).



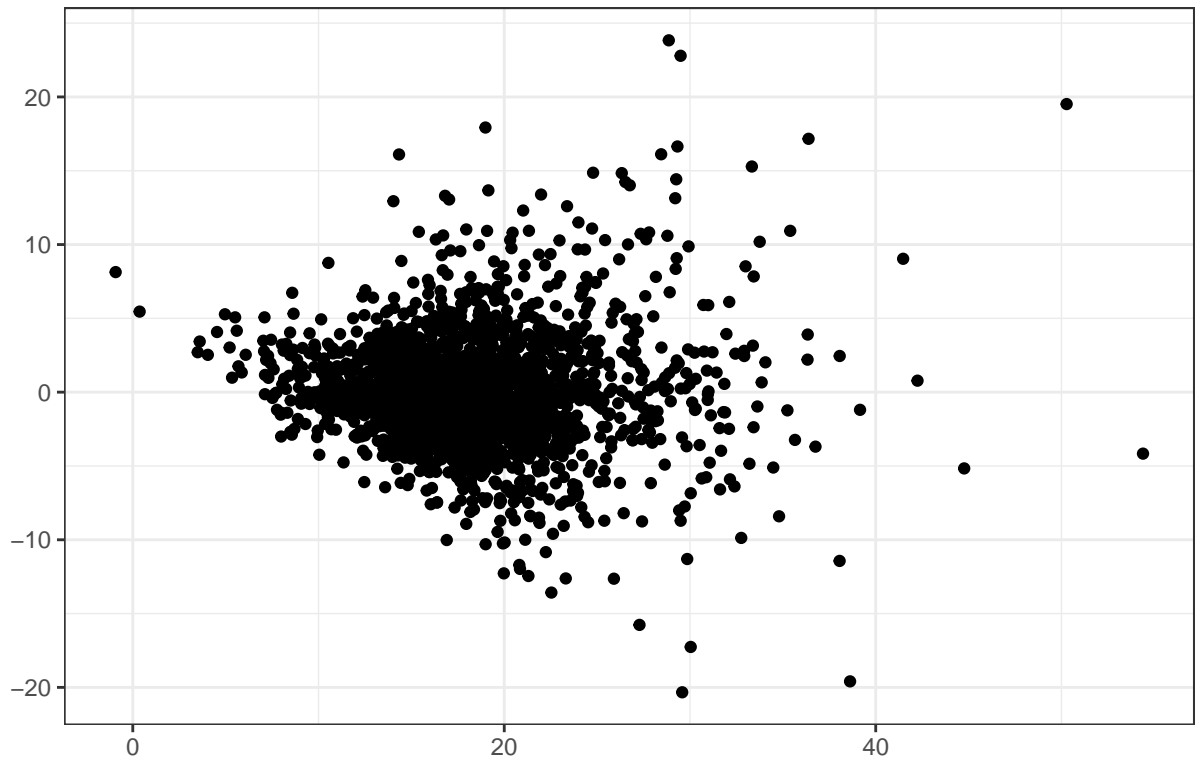
## Appendix E: Linear Model Diagnostics

```
qqnorm(resid(linmod))  
qqline(resid(linmod))
```



```
ggplot(mapping = aes(x = fitted(linmod),  
                      y = resid(linmod))) +  
  geom_point() +  
  theme_bw() +  
  labs(title = 'Residuals vs. Fitted Plot',  
        x = '',  
        y = '')
```

Residuals vs. Fitted Plot



## Appendix F: Transformed Linear Model Selection

```
# we create a new data frame for the transformed data
suicides_transformed <- suicides
# transform column then rename it
suicides_transformed$suicides <- sqrt(suicides_transformed$suicides)
suicides_transformed <- suicides_transformed %>%
  rename('sqrt_suicides' = 'suicides')
# do the same for the 7 predictors we are transforming
# uninsured
suicides_transformed$uninsured <- log(suicides_transformed$uninsured + 1)
suicides_transformed <- suicides_transformed %>%
  rename('log_uninsured' = 'uninsured')
# primary care physicians
suicides_transformed$primary_care_physicians <- log(suicides_transformed$primary_care_physicians + 1)
suicides_transformed <- suicides_transformed %>%
  rename('log_primary_care_physicians' = 'primary_care_physicians')
# mental health providers
suicides_transformed$mental_health_providers <- log(suicides_transformed$mental_health_providers + 1)
suicides_transformed <- suicides_transformed %>%
  rename('log_mental_health_providers' = 'mental_health_providers')
# unemployment
suicides_transformed$unemployment <- log(suicides_transformed$unemployment + 1)
suicides_transformed <- suicides_transformed %>%
  rename('log_unemployment' = 'unemployment')
# social associations
suicides_transformed$social_associations <- log(suicides_transformed$social_associations + 1)
suicides_transformed <- suicides_transformed %>%
  rename('log_social_associations' = 'social_associations')
# population
suicides_transformed$population <- log(suicides_transformed$population + 1)
suicides_transformed <- suicides_transformed %>%
  rename('log_population' = 'population')
# median income
suicides_transformed$median_income <- log(suicides_transformed$median_income + 1)
suicides_transformed <- suicides_transformed %>%
  rename('log_median_income' = 'median_income')

# perform model selection
names(suicides_transformed)[30] <- 'y'
suicides_transformed <- as.data.frame(suicides_transformed)
bestglm(na.omit(suicides_transformed), IC = 'AIC')$Subsets
```

```
##      (Intercept) food_environment_index access_to_exercise log_uninsured
## 0            TRUE                FALSE                FALSE        FALSE
## 1            TRUE                FALSE                FALSE        FALSE
## 2            TRUE                FALSE                FALSE        FALSE
## 3            TRUE                FALSE                FALSE        FALSE
## 4            TRUE                FALSE                FALSE         TRUE
## 5            TRUE                FALSE                FALSE         TRUE
## 6            TRUE                FALSE                FALSE         TRUE
## 7            TRUE                FALSE                FALSE         TRUE
## 8            TRUE                FALSE                FALSE         TRUE
```

## 9	TRUE	FALSE	FALSE	TRUE
## 10	TRUE	TRUE	FALSE	TRUE
## 11	TRUE	TRUE	FALSE	TRUE
## 12	TRUE	TRUE	FALSE	TRUE
## 13	TRUE	TRUE	FALSE	TRUE
## 14	TRUE	TRUE	FALSE	TRUE
## 15	TRUE	TRUE	FALSE	TRUE
## 16	TRUE	TRUE	FALSE	TRUE
## 17	TRUE	TRUE	FALSE	TRUE
## 18	TRUE	TRUE	FALSE	TRUE
## 19	TRUE	TRUE	TRUE	TRUE
## 20	TRUE	TRUE	TRUE	TRUE
## 21*	TRUE	TRUE	TRUE	TRUE
## 22	TRUE	TRUE	TRUE	TRUE
## 23	TRUE	TRUE	TRUE	TRUE
## 24	TRUE	TRUE	TRUE	TRUE
## 25	TRUE	TRUE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE	TRUE
##	log_primary_care_physicians	log_mental_health_providers		
## 0	FALSE	FALSE		
## 1	FALSE	FALSE		
## 2	FALSE	FALSE		
## 3	FALSE	FALSE		
## 4	FALSE	FALSE		
## 5	FALSE	FALSE		
## 6	FALSE	FALSE		
## 7	FALSE	FALSE		
## 8	FALSE	TRUE		
## 9	FALSE	TRUE		
## 10	FALSE	FALSE		
## 11	FALSE	TRUE		
## 12	FALSE	TRUE		
## 13	FALSE	TRUE		
## 14	FALSE	TRUE		
## 15	FALSE	TRUE		
## 16	FALSE	TRUE		
## 17	FALSE	TRUE		
## 18	FALSE	TRUE		
## 19	FALSE	TRUE		
## 20	FALSE	TRUE		
## 21*	FALSE	TRUE		
## 22	FALSE	TRUE		
## 23	FALSE	TRUE		
## 24	FALSE	TRUE		
## 25	FALSE	TRUE		
## 26	FALSE	TRUE		
## 27	FALSE	TRUE		
## 28	TRUE	TRUE		
## 29	TRUE	TRUE		
##	high_school_completion_rate	log_unemployment	income_inequality	
## 0	FALSE	FALSE	FALSE	

## 1	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE
## 4	FALSE	FALSE	FALSE
## 5	TRUE	FALSE	FALSE
## 6	TRUE	FALSE	FALSE
## 7	TRUE	FALSE	FALSE
## 8	TRUE	FALSE	FALSE
## 9	TRUE	FALSE	FALSE
## 10	TRUE	FALSE	FALSE
## 11	TRUE	FALSE	FALSE
## 12	TRUE	FALSE	FALSE
## 13	TRUE	FALSE	FALSE
## 14	TRUE	FALSE	FALSE
## 15	TRUE	FALSE	FALSE
## 16	TRUE	FALSE	FALSE
## 17	TRUE	FALSE	TRUE
## 18	TRUE	FALSE	TRUE
## 19	TRUE	FALSE	TRUE
## 20	TRUE	FALSE	TRUE
## 21*	TRUE	FALSE	TRUE
## 22	TRUE	FALSE	TRUE
## 23	TRUE	TRUE	TRUE
## 24	TRUE	TRUE	TRUE
## 25	TRUE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE
##	log_social_associations	air_pollution	drinking_water_violations
## 0	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE
## 4	FALSE	FALSE	FALSE
## 5	FALSE	FALSE	FALSE
## 6	FALSE	FALSE	FALSE
## 7	FALSE	FALSE	FALSE
## 8	FALSE	FALSE	FALSE
## 9	FALSE	FALSE	FALSE
## 10	FALSE	FALSE	FALSE
## 11	FALSE	FALSE	FALSE
## 12	FALSE	FALSE	FALSE
## 13	TRUE	FALSE	FALSE
## 14	TRUE	FALSE	FALSE
## 15	TRUE	TRUE	FALSE
## 16	TRUE	TRUE	FALSE
## 17	TRUE	FALSE	FALSE
## 18	TRUE	TRUE	FALSE
## 19	TRUE	TRUE	FALSE
## 20	TRUE	TRUE	FALSE
## 21*	TRUE	TRUE	TRUE
## 22	TRUE	TRUE	TRUE
## 23	TRUE	TRUE	TRUE

## 24	TRUE	TRUE	TRUE
## 25	TRUE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE
##	severe_housing_problems	food_insecurity	school_segregation
## 0	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE
## 4	FALSE	FALSE	FALSE
## 5	FALSE	FALSE	FALSE
## 6	FALSE	FALSE	FALSE
## 7	FALSE	FALSE	FALSE
## 8	FALSE	FALSE	FALSE
## 9	FALSE	FALSE	FALSE
## 10	FALSE	FALSE	FALSE
## 11	FALSE	FALSE	FALSE
## 12	FALSE	FALSE	FALSE
## 13	FALSE	FALSE	FALSE
## 14	FALSE	FALSE	FALSE
## 15	FALSE	FALSE	FALSE
## 16	TRUE	FALSE	FALSE
## 17	TRUE	FALSE	FALSE
## 18	TRUE	FALSE	FALSE
## 19	TRUE	FALSE	FALSE
## 20	TRUE	FALSE	FALSE
## 21*	TRUE	FALSE	FALSE
## 22	TRUE	FALSE	FALSE
## 23	TRUE	FALSE	FALSE
## 24	TRUE	TRUE	FALSE
## 25	TRUE	TRUE	FALSE
## 26	TRUE	TRUE	FALSE
## 27	TRUE	TRUE	FALSE
## 28	TRUE	TRUE	FALSE
## 29	TRUE	TRUE	TRUE
##	school_funding_adequacy	gender_pay_gap	
## 0	FALSE	FALSE	
## 1	FALSE	FALSE	
## 2	FALSE	FALSE	
## 3	FALSE	FALSE	
## 4	FALSE	FALSE	
## 5	FALSE	FALSE	
## 6	TRUE	FALSE	
## 7	FALSE	FALSE	
## 8	FALSE	FALSE	
## 9	TRUE	FALSE	
## 10	TRUE	FALSE	
## 11	TRUE	FALSE	
## 12	TRUE	FALSE	
## 13	TRUE	FALSE	
## 14	TRUE	FALSE	
## 15	TRUE	FALSE	

## 16	TRUE	FALSE			
## 17	TRUE	FALSE			
## 18	TRUE	FALSE			
## 19	TRUE	FALSE			
## 20	TRUE	FALSE			
## 21*	TRUE	FALSE			
## 22	TRUE	FALSE			
## 23	TRUE	FALSE			
## 24	TRUE	FALSE			
## 25	TRUE	FALSE			
## 26	TRUE	TRUE			
## 27	TRUE	TRUE			
## 28	TRUE	TRUE			
## 29	TRUE	TRUE			
##	nonwhite_white_residential_segregation		childcare_cost_burden		
## 0		FALSE		FALSE	
## 1		FALSE		FALSE	
## 2		FALSE		FALSE	
## 3		FALSE		FALSE	
## 4		FALSE		FALSE	
## 5		FALSE		FALSE	
## 6		FALSE		FALSE	
## 7		FALSE		FALSE	
## 8		FALSE		FALSE	
## 9		FALSE		FALSE	
## 10		TRUE		FALSE	
## 11		TRUE		FALSE	
## 12		TRUE		FALSE	
## 13		FALSE		FALSE	
## 14		TRUE		FALSE	
## 15		TRUE		FALSE	
## 16		TRUE		FALSE	
## 17		TRUE		FALSE	
## 18		TRUE		FALSE	
## 19		TRUE		FALSE	
## 20		TRUE		FALSE	
## 21*		TRUE		FALSE	
## 22		TRUE		FALSE	
## 23		TRUE		FALSE	
## 24		TRUE		FALSE	
## 25		TRUE		FALSE	
## 26		TRUE		FALSE	
## 27		TRUE		TRUE	
## 28		TRUE		TRUE	
## 29		TRUE		TRUE	
##	homeownership_rate	broadband_access	log_population	below_18	65_and_over
## 0	FALSE	FALSE	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE	FALSE	FALSE
## 2	FALSE	FALSE	FALSE	FALSE	FALSE
## 3	FALSE	FALSE	FALSE	FALSE	FALSE
## 4	FALSE	FALSE	FALSE	FALSE	FALSE
## 5	FALSE	FALSE	FALSE	FALSE	FALSE
## 6	FALSE	FALSE	FALSE	FALSE	FALSE
## 7	FALSE	FALSE	FALSE	TRUE	TRUE

## 8		FALSE	FALSE	FALSE	TRUE	TRUE
## 9		FALSE	FALSE	FALSE	TRUE	TRUE
## 10		FALSE	FALSE	FALSE	TRUE	TRUE
## 11		FALSE	FALSE	FALSE	TRUE	TRUE
## 12		FALSE	FALSE	FALSE	TRUE	TRUE
## 13		FALSE	FALSE	FALSE	TRUE	TRUE
## 14		FALSE	FALSE	FALSE	TRUE	TRUE
## 15		FALSE	FALSE	FALSE	TRUE	TRUE
## 16		FALSE	FALSE	FALSE	TRUE	TRUE
## 17		FALSE	FALSE	TRUE	TRUE	TRUE
## 18		FALSE	FALSE	TRUE	TRUE	TRUE
## 19		FALSE	FALSE	TRUE	TRUE	TRUE
## 20		FALSE	TRUE	TRUE	TRUE	TRUE
## 21*		FALSE	TRUE	TRUE	TRUE	TRUE
## 22		TRUE	TRUE	TRUE	TRUE	TRUE
## 23		TRUE	TRUE	TRUE	TRUE	TRUE
## 24		TRUE	TRUE	TRUE	TRUE	TRUE
## 25		TRUE	TRUE	TRUE	TRUE	TRUE
## 26		TRUE	TRUE	TRUE	TRUE	TRUE
## 27		TRUE	TRUE	TRUE	TRUE	TRUE
## 28		TRUE	TRUE	TRUE	TRUE	TRUE
## 29		TRUE	TRUE	TRUE	TRUE	TRUE
##	white_pop	female_pop	percent_rural	children_poverty	log_median_income	
## 0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## 1	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
## 2	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
## 3	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 4	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 5	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 6	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 7	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 8	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 9	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 10	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 11	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
## 12	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
## 13	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 14	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 15	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 16	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 17	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 18	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 19	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 20	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 21*	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 22	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 23	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 24	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
## 25	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
## 26	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
## 27	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
## 28	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
## 29	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
##	firearm_fatalities	logLikelihood		AIC		

```
## 0          FALSE      678.8234 -1357.647
## 1          TRUE      1116.1751 -2230.350
## 2          TRUE      1364.0530 -2724.106
## 3          TRUE      1443.0870 -2880.174
## 4          TRUE      1482.3056 -2956.611
## 5          TRUE      1537.4607 -3064.921
## 6          TRUE      1564.6617 -3117.323
## 7          TRUE      1616.0658 -3218.132
## 8          TRUE      1650.0866 -3284.173
## 9          TRUE      1670.3787 -3322.757
## 10         TRUE      1690.4307 -3360.861
## 11         TRUE      1710.3039 -3398.608
## 12         TRUE      1721.7527 -3419.505
## 13         TRUE      1730.8898 -3435.780
## 14         TRUE      1744.2676 -3460.535
## 15         TRUE      1748.5809 -3467.162
## 16         TRUE      1753.5645 -3475.129
## 17         TRUE      1757.7849 -3481.570
## 18         TRUE      1761.2021 -3486.404
## 19         TRUE      1763.3141 -3488.628
## 20         TRUE      1765.1359 -3490.272
## 21*        TRUE      1766.6856 -3491.371
## 22         TRUE      1767.6647 -3491.329
## 23         TRUE      1768.5266 -3491.053
## 24         TRUE      1769.2611 -3490.522
## 25         TRUE      1769.4858 -3488.972
## 26         TRUE      1769.5962 -3487.192
## 27         TRUE      1769.6875 -3485.375
## 28         TRUE      1769.6954 -3483.391
## 29         TRUE      1769.7030 -3481.406
```

```
linmod_transformed <- lm(y ~ food_environment_index + access_to_exercise + log_uninsured +
  log_mental_health_providers + high_school_completion_rate +
  income_inequality + log_social_associations + air_pollution +
  drinking_water_violations + severe_housing_problems +
  school_funding_adequacy + nonwhite_white_residential_segregation +
  broadband_access + log_population + below_18 + `65_and_over` + white_pop +
  female_pop + children_poverty + log_median_income + firearm_fatalities,
  data = suicides_transformed)
summary(linmod_transformed)
```

```
##
## Call:
## lm(formula = y ~ food_environment_index + access_to_exercise +
##     log_uninsured + log_mental_health_providers + high_school_completion_rate +
##     income_inequality + log_social_associations + air_pollution +
##     drinking_water_violations + severe_housing_problems + school_funding_adequacy +
##     nonwhite_white_residential_segregation + broadband_access +
##     log_population + below_18 + '65_and_over' + white_pop + female_pop +
##     children_poverty + log_median_income + firearm_fatalities,
##     data = suicides_transformed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```

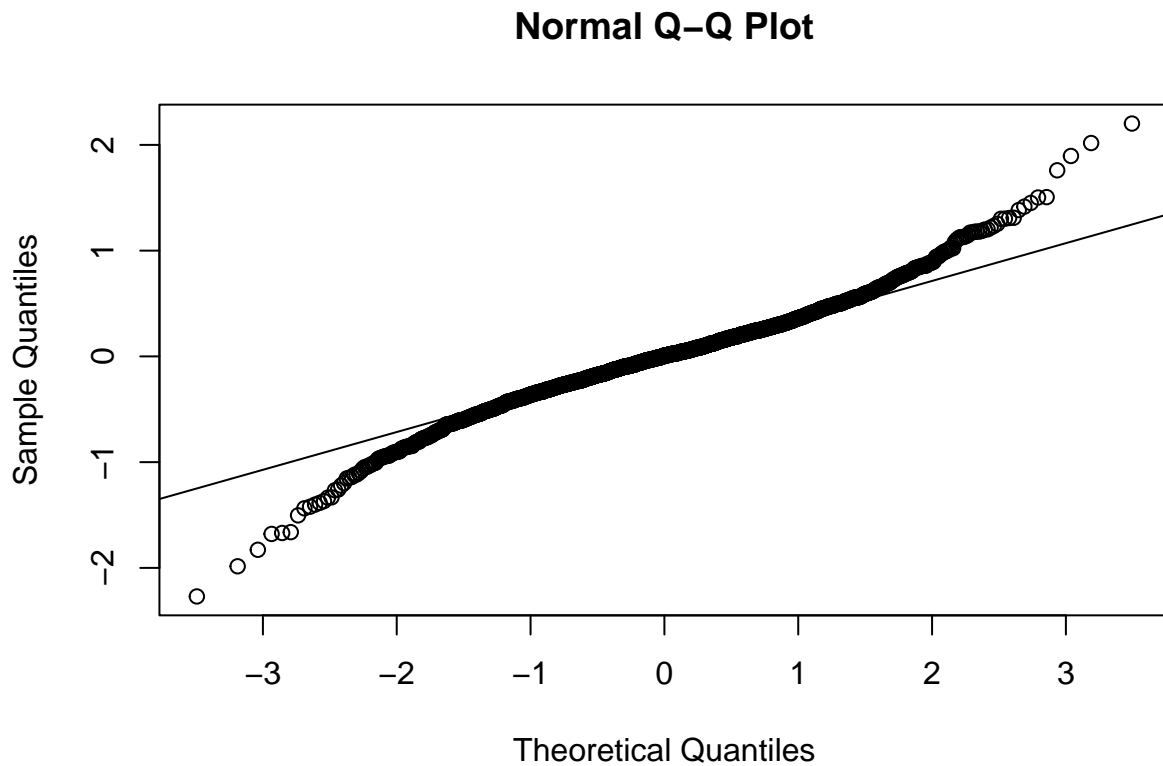
## -2.26995 -0.24180 0.00729 0.24005 2.20186
##
## Coefficients:
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.297e+00 1.111e+00 8.367 < 2e-16
## food_environment_index -1.129e-01 1.628e-02 -6.933 5.47e-12
## access_to_exercise 1.388e-01 6.221e-02 2.232 0.025739
## log_uninsured 2.849e+00 3.190e-01 8.930 < 2e-16
## log_mental_health_providers 3.191e+01 6.111e+00 5.221 1.95e-07
## high_school_completion_rate 2.225e+00 3.116e-01 7.140 1.28e-12
## income_inequality -6.188e-02 1.925e-02 -3.214 0.001330
## log_social_associations -1.314e-01 2.817e-02 -4.663 3.31e-06
## air_pollution -1.710e-02 6.360e-03 -2.689 0.007231
## drinking_water_violations 3.656e-02 2.001e-02 1.827 0.067844
## severe_housing_problems 1.530e+00 3.965e-01 3.860 0.000117
## school_funding_adequacy 1.925e-05 3.270e-06 5.886 4.60e-09
## nonwhite_white_residential_segregation -3.575e-03 9.405e-04 -3.801 0.000149
## broadband_access -4.166e-01 2.279e-01 -1.828 0.067623
## log_population -4.577e-02 1.447e-02 -3.163 0.001586
## below_18 5.020e+00 4.996e-01 10.048 < 2e-16
## '65_and_over' 3.771e+00 3.588e-01 10.510 < 2e-16
## white_pop 1.061e+00 9.229e-02 11.493 < 2e-16
## female_pop -7.538e+00 6.663e-01 -11.313 < 2e-16
## children_poverty -2.215e+00 3.243e-01 -6.830 1.11e-11
## log_median_income -4.170e-01 9.916e-02 -4.205 2.72e-05
## firearm_fatalities 6.066e-02 1.826e-03 33.230 < 2e-16
##
## (Intercept) ***
## food_environment_index ***
## access_to_exercise *
## log_uninsured ***
## log_mental_health_providers ***
## high_school_completion_rate ***
## income_inequality **
## log_social_associations ***
## air_pollution **
## drinking_water_violations .
## severe_housing_problems ***
## school_funding_adequacy ***
## nonwhite_white_residential_segregation ***
## broadband_access .
## log_population **
## below_18 ***
## '65_and_over' ***
## white_pop ***
## female_pop ***
## children_poverty ***
## log_median_income ***
## firearm_fatalities ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4287 on 2076 degrees of freedom
## (1044 observations deleted due to missingness)

```

```
## Multiple R-squared:  0.6527, Adjusted R-squared:  0.6492  
## F-statistic: 185.8 on 21 and 2076 DF,  p-value: < 2.2e-16
```

## Appendix G: Transformed Linear Model Diagnostics

```
qqnorm(resid(linmod_transformed))  
qqline(resid(linmod_transformed))
```



```
ggplot(mapping = aes(x = fitted(linmod_transformed),  
                      y = resid(linmod_transformed))) +  
  geom_point() +  
  theme_bw() +  
  labs(title = 'Residuals vs. Fitted Plot',  
        x = '',  
        y = '')
```

Residuals vs. Fitted Plot

