The Hygiene Service Industry's Impact on Economic Development and the Spread of Covid-19

Parv Joshi

Department of Economics, Stony Brook University

ECO 487: Independent Research in Economics & ECO 459: Write Effectively in Economics

Dr. María Pérez-Urdiales

May 10, 2021

Abstract

The coronavirus outbreak has drastically increased the demand for and use of hygiene products such as hand sanitizers and face masks. Several researchers have attempted to determine the effect of the hygiene product industry on the spread of the virus and how it may affect the GDP per capita of a country. This research is a preliminary study that focuses on the hygiene service industry instead of the hygiene product industry. This research examines if the hygiene service industry plays a significant role in economic development and the spread of the virus. Econometric analysis was performed using cross-country data on hygiene service measures collected by the *UNICEF*. The results indicate that some hygiene service factors like access to basic sanitation significantly predict economic development, while some factors like access to basic hygiene affect the spread of the virus. Also, some factors, like access to safe water, do not predict any of those.

Keywords: econometric analysis, coronavirus, hygiene service industry, economic development, instrumental variable regression

1. Introduction

When studying economic development, one often examines capital, labor, and technological growth. Modern economic development models also augment human capital, which accounts for the education and healthcare industries (Mankiw et al., 1992). However, one aspect people often fail to contemplate is the effect of the hygiene industry on economic development. Due to the coronavirus outbreak, the importance of the hygiene product industry has risen, and several studies have been conducted to determine its effect on the spread of the virus.

Before the development of vaccines, the only way to combat this disease was through lockdowns and hygiene measures. These measures caused a spike in the use of hygiene products such as hand sanitizers, disinfectants, tissues, face masks, disposable gloves, soaps, personal protective equipment (PPE) such as face shields, and immunity boosters such as vitamin-C supplements. In America alone, sales of hand sanitizers increased by 5,678%, cleaning products by 344%, soap by 170%, tissues by 122%, and immunity boosters by 88% (EnsembleIQ, 2020; Statista, 2020).

I studied the effect of the hygiene service industry instead of the hygiene product industry on economic development and in controlling the spread of the coronavirus. This research was a preliminary study where I used pooled data econometric analysis to answer the following research questions:

1. Do hygiene service factors significantly affect the spread of the coronavirus?
2. Do hygiene service factors significantly affect a country’s standard of living?

The first question is extremely relevant at this time, as it would help us better understand which hygiene service measures, if any, are most effective in combating the current pandemic. Moreover, it can help us know which hygiene services must be improved to prevent such pandemics in the future. The second question is also relevant since it aims to find which hygiene services contribute to economic growth.

Previously, some researchers have studied the role of the hygiene industry on economic development, while some have studied the role of the standard of living on the hygiene industry. To understand which variable explains the other, I studied both directions of causation: the possibility that hygiene services affect a country's standard of living and the possibility that standard of living affects hygiene services. I used instrumental variable (IV) regression to correct for simultaneous causality in the single-directional model I considered. My single-directional used the number of Covid-19 positive cases and standard of living as the response variables for my research questions, respectively.

For this study, I used three hygiene service factors: access to basic sanitation, hygiene, and safe water. UNICEF collected data for those variables in 2020. For the first research question, the average number of coronavirus cases in a country was used as the response variable, while for the second question, the standard of living was used. Data for both those variables were collected from the *Master Covid-19 Dataset* created by Appel et al. (2021). Other variables related to population, healthcare, and the food industry were also considered while performing the analysis.

I used pooled data analysis since data for all variables I considered were not collected at once; instead, they were compiled from four different sources: *The World Bank*, *UNICEF*, *Master Covid-19 Dataset*, and *Our World in Data*. I used econometrics analysis for the study since it aimed to establish causality and find its direction. For example, it helped determine whether a country's hygiene services affect its standard of living or whether it is the standard of living that affects the hygiene services.

The results revealed that increased access to basic sanitation significantly improves economic development. In contrast, increased access to basic hygiene helps reduce the spread of the coronavirus, and access to safe water was not a significant predictor of both.

In the next section, I will present a literature review of past studies conducted on this topic. In Section 3, I will detail the methodology used, and in Section 4, I will present the results of my research and discuss its policy implications. Lastly, I will present my conclusions in the final section of this paper.

2. Literature Review

Many studies have been conducted to explain the effect of the hygiene industry on a country's standard of living and the spread of Covid-19. As mentioned in Section 1, different studies focused on different directions of causation. Hence, the primary focus of this review was to understand the direction of causation better and discover the variables relevant to this study. Let us first understand what factors may be relevant to predict the spread of Covid-19.

One of the first cross-country studies on how the coronavirus affected the hygiene product industry was conducted by Orion Market Research Private Limited in March 2020. They used consumer behavior, corporate data, micro- and macro-economic factors, laws, and government policies to develop insights on the increased use of hygiene products due to the pandemic, which they compiled in a report (Research and Markets Ltd., 2020).

Moore et al. (2021) found that "hand hygiene performance (HHP) rates" in nine hospitals increased from 46% to 60% in April 2020 and then declined to 56% in May 2020 at a 1% level of significance. They concluded that "HHP shifted in multiple directions during the early stages of the pandemic," which may have been due to the "increased emphasis on the importance of hand hygiene," government policies mandating the use of face masks, and school and workplace closures (Moore et al., 2021, pp. 30-32). This suggested the relevance of variables relating to government policies. Moreover, it suggested that hygiene measures may be the dependent variable since the spread of Covid-19 affects government policies, which in turn affects the hygiene measures.

The American Cleaning Institute (2020) released a report that addressed the impact of Covid-19 on the cleaning product industry. They claimed that hand sanitizer production increased by 427,000 gallons globally, causing a 23% increase in the average “production volume of sanitizing products” (American Cleaning Institute, 2020). Meanwhile, its sales increased by 5,678% (EnsembleIQ, 2020). Although the effect of hand sanitizers does not directly relate to my study, increased use of sanitizers suggested increased use of hygiene services measures.

Not only sanitizers, but the entire hygiene industry had a similar trend. According to Global Industry Analysts Inc (2020), the “global market for hand hygiene” reached $1.2 billion in 2020, which is projected to increase to $2.3 billion by 2027, with an average “compound annual growth rate (CAGR) of 9.6%.” Considering the hygiene industry as a whole, it was unpredictable whether the increase was only due to the hygiene product industry or the hygiene service industry also had a significant role.

Although some hygiene products’ demand and sales had increased due to the pandemic, some hygiene services experienced adverse effects. According to Butler et al. (2020), the demand for “industrial water” would decline by 27%, resulting in a revenue loss for the water utility industry across the globe and disrupting “the whole water supply chain” (Butler et al., 2020). They suspect that the disruption was caused due to decreased supply during factory closures (Butler et al., 2020). This implied that access to safe water was a hygiene service factor relevant to this study.

According to the American Society for Parenteral and Enteral Nutrition (2020), increased access to water and frequent rehydration helps in clearing “respiratory secretions” from the lungs, which help in reducing the effect of the coronavirus. Moreover, according to a research by Donde et al. (2021), poor sanitation and waste management had "[a] significantly higher prevalence of [the] SARS-CoV-2 infection." Hence, access to water, sanitation services, and waste management may be significant predictors of the spread of Covid-19.

Furthermore, there were some hygiene products whose demand and sales decreased. According to a report released by Euromonitor International (2020), the demand for away-from-home tissue products "dropped significantly in 2020 due to business and travel disruptions," although their sales substantially increased in the first two quarters. This suggested that considering other factors in the analysis was essential.

Let us now explore the literature that aims to discover which hygiene service factors may be relevant in explaining economic development: In 2011, Van Minh and Hung (2011) explored the "economic aspects of sanitation in developing countries," where they concluded that "sanitation is undeniably a profitable investment" since it "not only saves lives but also provides a foundation for economic growth," suggesting it had a positive correlation with economic development. Moreover, according to a 2007 article by WaterAid America, improvement in sanitation services establish a foundation for economic growth in addition to saving lives, promoting gender equality, and encouraging education (Kemeny, 2007, "Conclusion"). These suggested that sanitation was an essential measure of the hygiene service industry. Additionally, educational and environmental factors may be relevant predictors of a country's standard of living.

Investment in hygiene measures is crucial for economic development, especially in developing countries. As noted in a 2017 report by Hygiene Matters, "[a]n average of 6,500 Indian Rupees per person (about USD 97) was lost in India annually due to a lack of cleanliness and hygiene" (World Health Organization, 2017, as cited in Hygiene Matters, 2017). Moreover, their 2017 survey results claim that about 42% of people in developing countries refrain from using a public toilet due to hygiene concerns (Hygiene Matters, 2017). This suggested that investments in the hygiene service industry would help reduce a massive cost for developing countries by improving their long-term standard of living. Therefore, variables relating to hygiene measures were crucial for the analysis.

Apart from proper sanitation and hygiene, access to safe water is also crucial. According to Sustainable Sanitation and Water Management (SSWM), the "[e]conomic benefits of improved water and sanitation services" include less "direct and indirect health costs," increased "return on investments in education," higher productivity, reduced "investments in improved water supply," less economic harm due to polluted water and sanitation, and increased "tourism revenues," which all increase a country's standard of living (Keller, 2019). This indicated the importance of safe water accessibility to this research.

This review aimed to provide different perspectives used by past researchers on the topic that I focus on in this paper. The review's significance lied in forming the hypotheses, deciding which variables to consider, and predicting the results of this study. I found that the most relevant variables to consider were access to basic hygiene, sanitation, and water services. Moreover, this review suggested that hygiene services were most likely to be the explanatory variables.

Although much discussion has been conducted on this topic, most of the sources I found were survey reports or articles, and econometric analysis was rarely used for such a study. Hence, it was vital to conduct more research using various methods to better understand the hygiene service industry's effect on economic development and the spread of the coronavirus.

3. Methodology

3.1. Empirical Framework

To accomplish my research objectives, I first formulated my hypotheses as the following:

1. More hygiene measures should contribute to reducing the spread of the coronavirus. Therefore, countries with a more developed hygiene industry may have a lower infection rate and, as a consequence, fewer cases.
2. Hygiene services usually impinge on healthcare services, which play a huge role in economic development. Therefore, the hygiene services may contribute to an increment in a country's long-term standard of living.

To test these hypotheses, I formulated the empirical framework by creating equations for each of my research questions. The following sub-sections detail the equations of the basic models I used for each of my research questions. Model specifications were not included in these basic models since they were computed after finding the best dependent and independent variable transformations. Precise equations for each specification are summarized in Section 4.

3.1.1. The Effect of the Hygiene Service Industry on the Spread of Covid-19

For this model, my objective was to find the effect of hygiene, water, and sanitation services on the average number of Covid-19 cases in a country. However, I wanted to regress on all other variables as well, to check whether those variables would still be significant despite the inclusion of the other variables. Hence, the first model I tested was:

$$CASES = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times SL) + (\beta_8 \times HAQI) + \epsilon \qquad (1)$$

where CASES is the average number of positive coronavirus cases, HCI is the human capital index, MEAT is the meat consumption per capita, PD is the population density, HYGIENE is the percentage of the people with access to basic hygiene services, SANITATION is the percentage of the people with access to basic sanitation services, WATER is the percentage of the people with access to safe drinking water, SL is the standard of living, and HAQI is the health access and quality index.

3.1.2. The Effect of the Hygiene Service Industry on the Standard of Living

This model resembled the previous model, but the response variable here was the standard of living instead of the average number of Covid-19 cases in a country. Hence, the second model I tested was:

$$SL = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times CASES) + (\beta_8 \times HAQI) + \epsilon \quad (2)$$

where SL is the standard of living, HCI is the human capital index, MEAT is the meat consumption per capita, PD is the population density, HYGIENE is the percentage of the people with access to basic hygiene services, SANITATION is the percentage of the people with access to basic sanitation services, WATER is the percentage of the people with access to safe drinking water, CASES is the average number of positive coronavirus cases, and HAQI is the health access and quality index.

3.2. Data Collection

After formulating the research question, I considered nine variables for the study. I found 70 observations in total, all of which were country-wise, none of them being categorical. One caveat of the data collection process was that although I considered the most recent data, all the variables were not observed in the same year. However, it was expected that the data would be reasonably consistent over time. Moreover, data for only 70 countries could be collected. The following sub-sections detail these variables, as well as their collection procedures. Documentation of the data is summarized in Appendix A.

3.2.1. CASES

The variable CASES measured the average number of new coronavirus-positive cases for a country on a seven-day rolling basis as of March 1, 2021 (Appel et al., 2021). Data for this variable was gathered from the *Master Covid-19 Dataset* by Appel et al. (2021). The data was developed based on a seven-day rolling basis to create a standard unit of measurement since some countries only released their data weekly (Appel et al., 2021). This variable was the response variable for the model in my first research question.

3.2.2. HCI

The Human Capital Index (HCI) was considered for the study since it takes into account the education and healthcare industries and is used to explain the standard of living in the augmented Solow model (Mankiw et al., 1992). This index uses values from 0 to 100 to represent the average levels of education and health of an average individual in a country (Mankiw et al., 1992). It is found by taking the aggregate mean of education and health measures (Kraay, 2019).

Data of the year 2016 (most recent available) was collected from *The World Bank's* online database (The World Bank, 2021b). In Section 2, I found that factors relating to education and healthcare may affect both our response variables; hence, HCI was an important variable for this study.

3.2.3. MEAT

The variable MEAT measures the meat consumption per capita of a country in kilograms (Ritchie & Roser, 2019). Data of the year 2019 (most recent available) was collected from *Our World in Data's* online database (Ritchie & Roser, 2019). This variable from the food industry is considered a hygiene factor in many countries, and its inclusion could help to check if it affects the spread of the coronavirus and a country's standard of living.

Although this variable may not directly correlate with any of my response variables, it can help to account for spurious correlations. In early 2020, it was rumored that a non-meat-based diet could "prevent COVID-19 infections;" however, there was no evidence of it (Xinhua, 2020). The use of this variable could help determine if the rumor was true or not.

3.2.4. PD

Population density (PD) was used since it might have significant correlations with the spread of the coronavirus, the standards of the hygiene service industry, and the standard of living of a country. Data for this variable was also collected from the *Master Covid-19 Dataset* by Appel et al. (2021). Population density is the "[n]umber of people divided by land area, measured in square kilometers," and the data of the year 2019 (most recent available) was collected (Appel et al., 2021).

3.2.5. HYGIENE

HYGIENE is the first of the three hygiene service measures I considered for this study. Hence, it was one of the main explanatory variables in this study since my research questions aim to identify its effect on the two response variables. Data for this variable was gathered by *UNICEF* in 2020, and it measures the percentage of the population of a country with access to at least basic hygiene services (United Nations International Children's Emergency Fund, 2020).

3.2.6. SANITATION

SANITATION measures the percentage of the population of a country that has access to at least basic sanitation services, and its data was collected by *UNICEF* in 2020 (United Nations International Children's Emergency Fund, 2020). This variable was the second of the three proxies used to measure the hygiene service industry, and hence was one of my main explanatory variables.

3.2.7. WATER

WATER, also one of my main explanatory variables, was the last hygiene service measure I considered for this study. *UNICEF* collected its data in 2020, and it measures the percentage of the population of a country with access to safe water for drinking (United Nations International Children's Emergency Fund, 2020).

3.2.8. SL

The standard of living (SL) was used to measure economic development. Data of the year 2018 (most recent available) was collected from *The World Bank's* online database (The World Bank, 2018). The standard of living was purchasing power parity (PPP) adjusted for current U.S. dollars (The World Bank, 2018).

It was measured by first summing the "gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products," and dividing it "by [the] midyear population" (The World Bank, 2021a). It is the response variable for the model in my second research question.

3.2.9. HAQI

The Health Access and Quality Index (HAQI) was used to measure the quality and accessibility of healthcare institutions in a country. This variable ranged from 0 to 100 and was "based on death rates from 32 causes of death that could be avoided by timely and effective medical care" (Global Change Data Lab, 2016). Data of the year 2016 (most recent available) was collected from *Our World in Data's* online database (Global Change Data Lab, 2016). Since it is a variable related to healthcare, it may significantly affect both response variables.

3.3. Descriptive Statistics and Data Visualization

Table 1 provides the summary statistics for all the variables.

Table 1: Summary Statistics

Variable	Mean	Minimum	1st Quartile	Median	3rd Quartile	Maximum
CASES	877.329	0.00	19.71	82.57	608.61	20240.43
HCI	47.15	30	39.70	45.20	53.33	75.60
MEAT	26.91	3.78	13.07	20.29	36.30	88.37
PD	144.10	1.98	34.93	77.95	145.14	1454.43
HYGIENE	48.107	1.188	23.006	44.296	78.631	98.999

SANITATION	55.887	7.316	33.663	55.935	87.016	99.000
WATER	79.13	38.70	66.47	81.32	94.14	99.00
SL	2997.6	411.6	1082.8	2023.1	4301.7	10626.5
HAQI	51.57	32.50	43.67	49.75	59.88	75.50

From the table, one could observe that CASES took the minimum value of zero. Therefore, using the logarithmic transformation of CASES was not possible since the logarithm of zero is not defined. All the other variables were strictly positive, and hence their logarithmic transformations could be used.

To better understand the relationship between the variables, I computed the Pearson correlations for each variable. Table 2 provides the correlation matrix for this data.

Table 2: Correlation Matrix

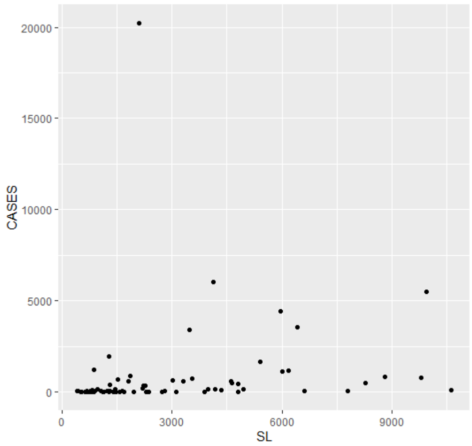
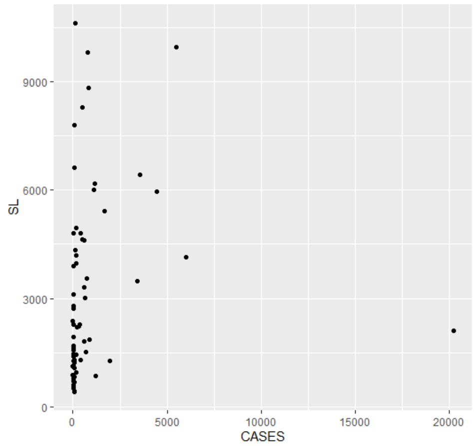
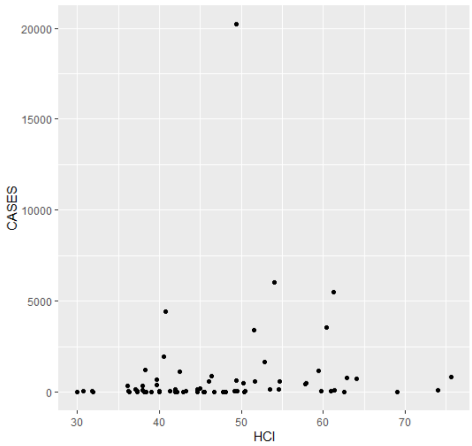
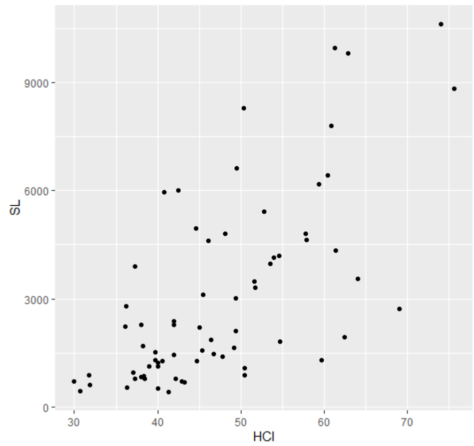
Variable	CASES	HCI	MEAT	PD	HYGIENE	SANITATION	WATER	SL	HAQI
CASES	1.00	0.13	-0.04	0.15	0.19	0.17	0.25	0.15	0.04
HCI	0.13	1.00	0.54	0.25	0.69	0.69	0.69	0.68	0.78
MEAT	-0.04	0.54	1.00	-0.24	0.50	0.51	0.52	0.63	0.52
PD	0.15	0.25	-0.24	1.00	0.09	0.14	0.21	0.16	0.20
HYGIENE	0.19	0.69	0.50	0.09	1.00	0.87	0.79	0.68	0.79
SANITATION	0.17	0.69	0.51	0.14	0.87	1.00	0.81	0.71	0.82
WATER	0.25	0.69	0.52	0.21	0.79	0.81	1.00	0.68	0.74
SL	0.15	0.68	0.63	0.16	0.68	0.71	0.68	1.00	0.72
HAQI	0.04	0.78	0.52	0.20	0.79	0.82	0.74	0.72	1.00

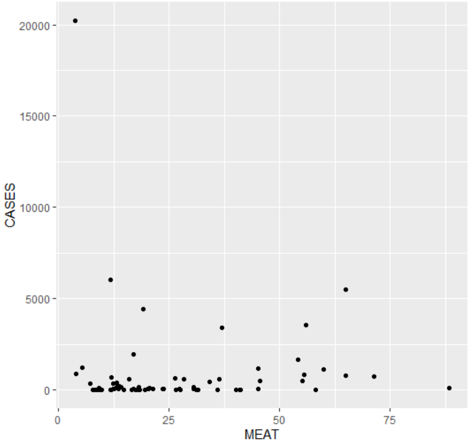
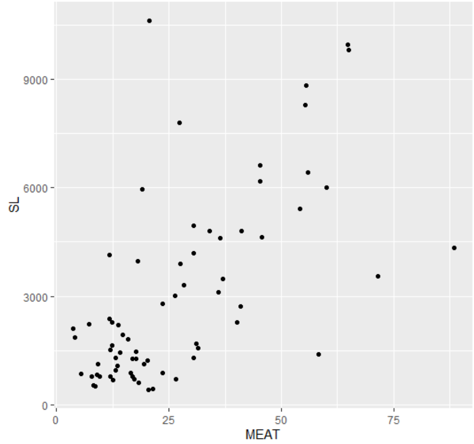
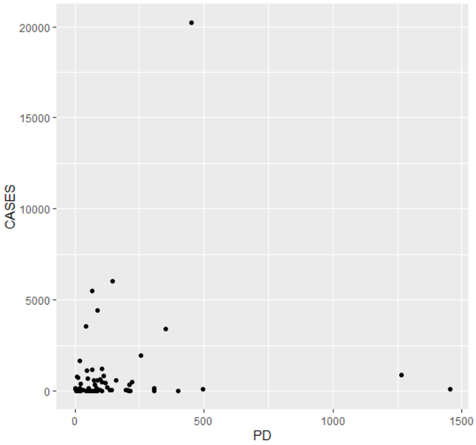
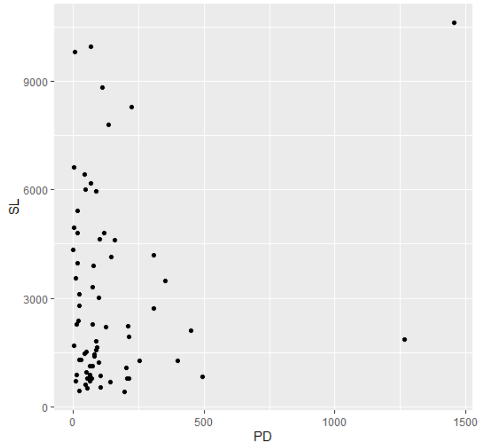
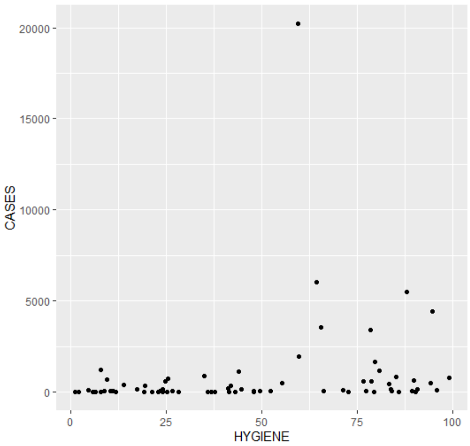
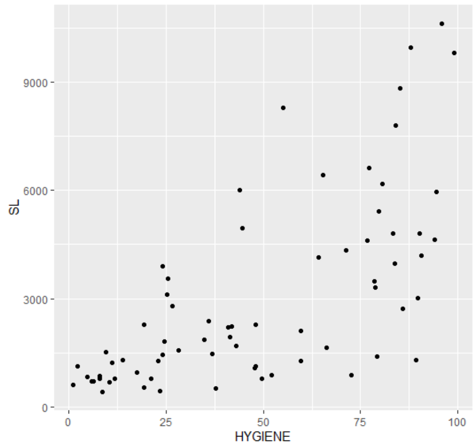
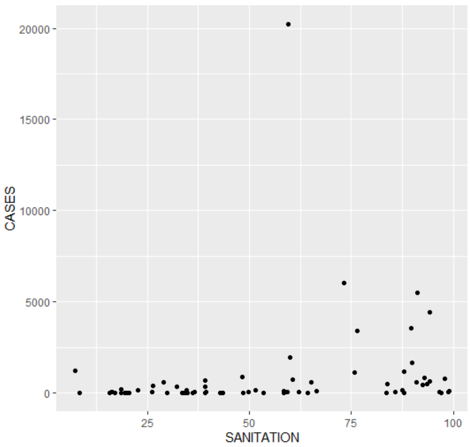
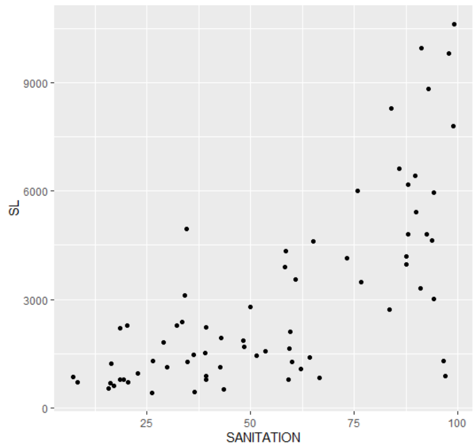
From the table, one could notice that CASES and PD had very low correlations with all other variables. Hence, HYGIENE, SANITATION, and WATER were not likely to be significant predictors of CASES. Apart from this, one could observe that HCI, HYGIENE, SANITATION, WATER, SL, and HAQI had strong correlations with each other. This suggested that the hygiene service industry may significantly affect economic development.

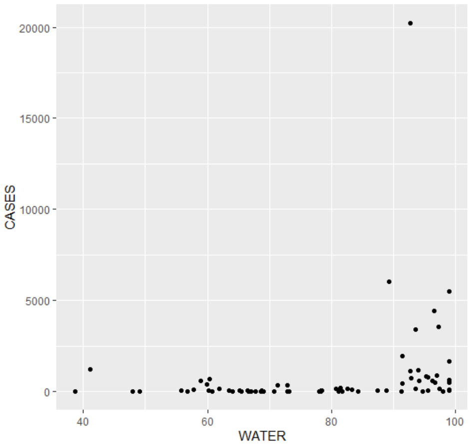
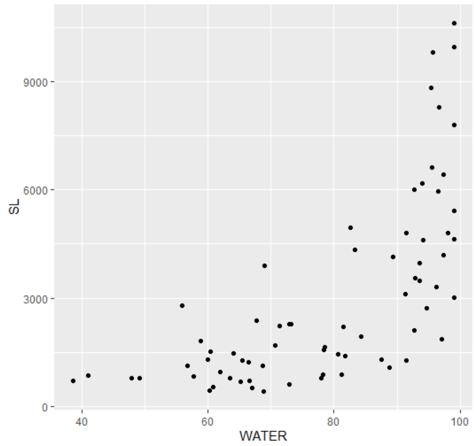
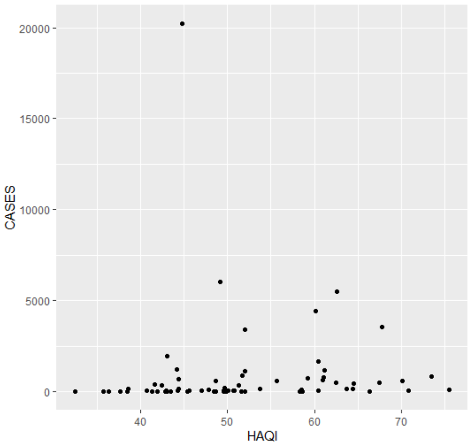
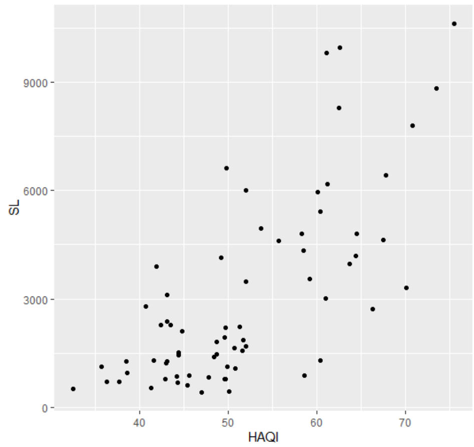
Correlations of MEAT with other variables were primarily moderate. Hence, its significance to this study could not be predicted. Moreover, it was interesting to note that MEAT's correlations with CASES and PD were small but negative. That suggested that low meat consumption could lead to fewer Covid-19 cases.

Apart from examining the summary statistics and the Pearson correlations, I also drew inferences by visualizing the data. I observed the scatterplots of each variable with both my responses to make inferences from them. Table 3 illustrates the scatterplots.

Table 3: Scatterplots

X\Y	**CASES**	**SL**
SL/CASES		
HCI		

MEAT		
PD		
HYGIENE		
SANITATION		

WATER		
HAQI		

Consistent with my previous predictions, I observed that CASES and PD had slopes close to zero in plots with most other variables. That suggested that my first research question model might not be very strong since it used CASES as the response variable. Moreover, HCI, HYGIENE, SANITATION, WATER, SL, and HAQI had positive-sloping graphs that indicated a strong relationship between the variables. Again, MEAT had points widely scattered in graphs for almost all variables, and hence its behavior was unpredictable.

3.4. Econometric Concepts

I primarily used the method of Ordinary Least Squares (OLS) regression to check for causality among the variables, as it aimed to minimize the sum of squared residuals by estimating the parameters of the model (Stock & Watson, 2008). However, such a model could be associated with endogeneity bias. Endogeneity bias is the error that could be caused due to omitted variables, error in measurement, and simultaneous causality (Stock & Watson, 2008).

As noted in Section 2, environmental factors may have been significant for the study; hence, its exclusion may have contributed to the endogeneity. Measurement error is not common for macroeconomic data, especially when the data is collected from the databases of reliable sources such as *The World Bank* and *UNICEF*. However, as addressed already, our observations for different variables were not of the same year. That may have contributed to the endogeneity due to error in measurement. Lastly, the possibility of bidirectional causality could not be ruled out, causing a bias in my single-directional model.

To address endogeneity bias, I used instrumental variable (IV) regression, which allowed me to find and account for spurious correlations. This type of regression required instrumental variables (also known as instruments) that formed a two-staged regression model. Finding valid instruments are usually difficult since they must follow two conditions:

1. Instrument relevance – there must be a significant correlation between the independent variables and the instrumental variables, and
2. Instrument exogeneity – instrumental variables must not be correlated with the error term.

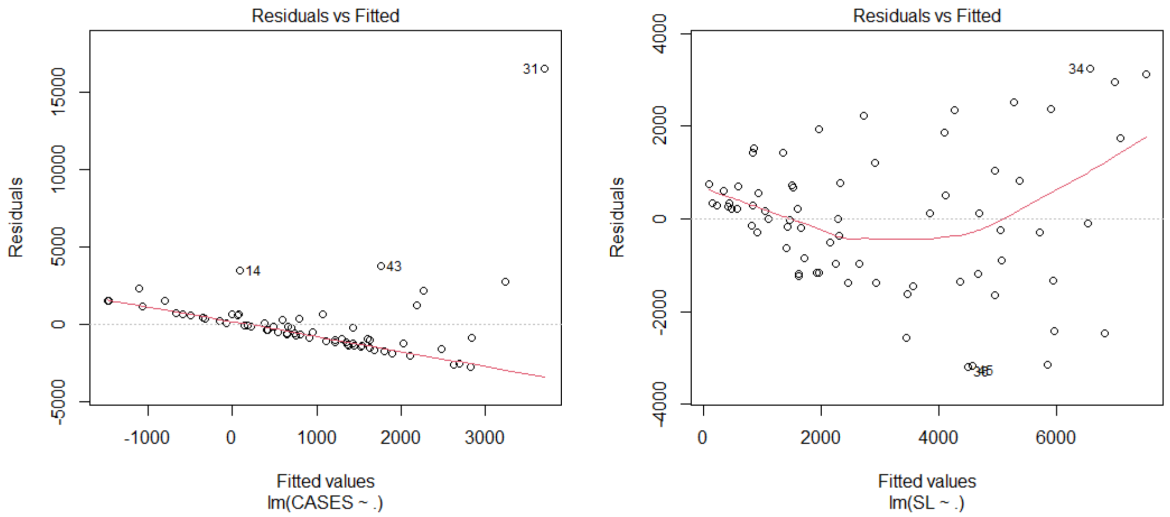
Moreover, a valid instrument must not directly affect the response variable (Stock & Watson, 2008). In order to perform IV regression, I first tested these assumptions, as detailed in Section 4.

In addition to this, I used model specifications such as including interaction terms, using variable transformations, and correcting for heteroskedasticity robust standard errors. I used linear, quadratic, cubic, and logarithmic transformations for independent variables, while for response variables, I used the logarithmic and the Box-Cox power transformations. The Box-

Cox transformation could help formulate a “valid distribution-free procedure” for the study (Sakia, 1992, p. 1).

It was important to correct my model for heteroskedasticity since non-homogenous variances make an OLS estimator biased, as the method of OLS assumes homoskedasticity (Stock & Watson, 2008). Models with several variables and transformations usually have non-homogeneous variances, and hence it was a good practice to correct for heteroskedasticity. For models one and two, the presence of heteroskedasticity could be confirmed by plotting the residuals versus fitted plots, as shown in Figure 1.

Figure 1: Residuals vs. Fitted Plots for Models 1 and 2



One could notice from the residuals versus fitted plots that the points were not evenly distributed along the zero-residual line. That suggested that the variances were non-homogenous, and hence there was a need to correct for heteroskedasticity. Since homoskedasticity did not hold for our basic models, it was important to correct for heteroskedasticity for all models to obtain unbiased estimates and standard errors.

4. Results and Policy Implications

4.1. Results

For both the basic equations I created, I first found the best transformations for each independent variable with respect to its response variable. As mentioned earlier, I performed linear, quadratic, cubic, and logarithmic transformations for the independent variables. I used Wald tests to compare the polynomial transformations, while to compare the best polynomial transformation with the logarithmic transformation, I plotted the fitted models to check which transformation best explained the variation of the data. Details of the computation of the best independent variable transformations for equations 1 and 2 can be found in Appendix B and C, respectively.

Using those transformations with different combinations of interaction terms and response variable transformations, I computed several models. Equations including the specifications of the models are detailed in Appendix D. Details of the computation, summary results, and the heteroskedasticity robust standard errors of all those models for equations 1 and 2 can be found in Appendix E and F, respectively. Comparing those models using goodness of fit measures, I found the best models for equations 1 and 2, which are summarized in Table 4.

Table 4: Significant Variables and their Estimates

(Equation) **Regressor \ Response**	**(1)** **CASES**	**(2)** **SL**
HYGIENE	– 45.883 ** (– 2.2597)	61.279 ** (2.1516)
PD	71.403 *** (3.2794)	2.078 (0.3769)
CASES		– 0.85007 *** (– 2.8767)
I(HAQI^3)		0.26017 * (1.6929)
SANITATION	– 293.59 ** (– 2.1763)	79.840 (0.6315)

I(SANITATION^2)	5.8435 ** (2.0862)	– 2.039 (– 0.7068)
I(SANITATION^3)	– 0.035285 ** (– 2.0364)	0.018 (0.9831)
MEAT:PD	– 0.67434 *** (– 2.7375)	
HYGIENE:PD	0.57993 *** (3.3398)	
PD:SANITATION	0.41591 ** (2.6270)	
HAQI:PD	– 2.0875 *** (– 3.2411)	
HYGIENE:CASES		0.015612 *** (3.2095)
MEAT:SANITATION		1.8647 ** (2.2056)
HYGIENE:SANITATION		– 1.2891 ** (– 2.0376)
Constant	– 8465.6 (– 0.5305)	– 29506 (– 1.2059)
Observations	70	70

* (p < 0.1), ** (p < 0.05), *** (p < 0.01); t-statistic in parentheses.

The multiple R^2 and the adjusted R^2 values for model one were 0.7339 and 0.6328, respectively, while for model two, they were 0.8386 and 0.7679. This suggested that model one collectively explained 63.28%, and model two explained 76.79% of the variance in the data. Therefore, both these models were strong and fit most of the observed data.

However, as discussed already, these results were subject to endogeneity bias. For this reason, I performed IV regression. For both models, my endogenous variables were the best transformations of HYGIENE, SANITATION, and WATER. For those endogenous variables, I first decided which variables to use as instruments and tested their assumptions.

For the first model, I considered HCI, HAQI, MEAT, and log(SL). It was unlikely that any of these were correlated to the error term since all such variables were used as endogenous variables in the OLS model, except for a variable that would measure the effect of vaccines.

That exception may have affected the correlation of HCI and HAQI with the error term. However, since the data was collected before the development of vaccines, one can assume that such a correlation may be negligible. Hence, all four variables satisfied the assumption of instrument exogeneity.

HCI accounts for an average individual's education and health, while HAQI accounts for the quality and accessibility of a country's healthcare system. Therefore, both could correlate to hygiene service measures and the spread of Covid-19. Hence, they satisfy the assumption of instrument relevance. Meat consumption and the standard of living may also influence hygiene services. Hence MEAT and log(SL) also satisfy the assumption of instrument relevance. This claim was confirmed by computing the second-staged model. Details of its computation can be found in Appendix G.

Thus, for the IV regression model of the first research question, I used HYGIENE^3, SANITATION^3, and WATER^3 as the endogenous variables; MEAT and log(SL) as the instruments; and PD, MEAT*PD, MEAT*HAQI, PD*HYGIENE, PD*SANITATION, PD*HAQI, and WATER*SL as the control variables.

For the second model, I considered MEAT^3, PD^3, and HCI. Population density could correlate to many factors not included in the model, such as immigration status and climate. Also, the human capital index may correlate to the quality of education. These factors could be a part of the error term since they may affect a country's standard of living. Hence, PD^3 and HCI could significantly be correlated with the error term, violating the assumption of instrument exogeneity. Meat consumption could not affect any other factor excluded from the OLS model, and hence MEAT^3 satisfied the assumption.

It was not intuitive if meat consumption affected hygiene services or not, so it was questionable if MEAT satisfied the assumption of instrument relevance. However, the OLS estimates from Table 3 showed that the interaction term of MEAT and SANITATION was significant at a 5% level of significance. Therefore, I computed the second-staged model for the same. Details of its computation can be found in Appendix H. I found that $MEAT^3$ was significant but had an estimate close to zero. Hence, $MEAT^3$ was a weak instrument.

Thus, for the IV regression model of the second research question, I used $HYGIENE^3$, $SANITATION^3$, and $WATER^3$ as the endogenous variables; $MEAT^3$ as the instrument; and $HAQI^3$, CASES*HYGIENE, MEAT*SANITATION, and HYGIENE*SANITATION as the control variables.

Using those instruments, I computed the heteroskedasticity robust standard errors for both models using IV regression. Details of their computation can be found in Appendix I. The results for the same are summarized in Table 5.

Table 5: IV Regression Results

(Equation) Regressor \ Response	**(1) CASES**	**(2) SL**
HYGIENE	– 45.192 ** (– 2.509)	70.418 (1.604)
PD	72.655 *** (3.250)	
HAQI	95.316 (1.448)	3,423.715 ** (2.427)
I(HAQI^2)		– 66.464 ** (– 2.389)
I(HAQI^3)		0.410 ** (2.329)
SANITATION	– 70.540 (– 0.041)	1,261.738 ** (2.280)
I(SANITATION^2)	1.051 (0.028)	– 27.024 ** (– 2.341)
I(SANITATION^3)	– 0.006 (– 0.027)	0.172 ** (2.419)
HYGIENE:PD	0.566 *** (3.112)	
PD:HAQI	– 2.133 *** (– 2.933)	

HYGIENE:SANITATION		– 1.474 * (– 1.819)
Constant	2,143.040 (0.060)	– 14,638.580 (– 0.155)
Observations	70	70

* (p < 0.1), ** (p < 0.05), *** (p < 0.01); z-statistic in parentheses.

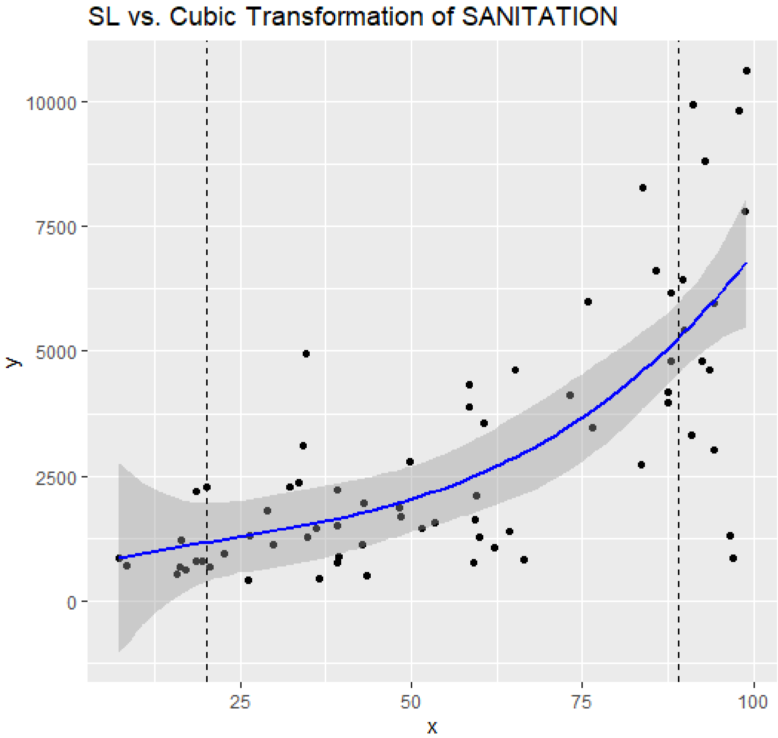
4.2. Implications on Economic Policy

From the results of equation one, it was found that HYGIENE had a negative estimate and was significant at a 5% level of significance. Therefore, access to basic hygiene services plays a significant role in reducing the spread of the coronavirus. Similarly, PD had a positive estimate and was significant at a 1% level of significance, indicating that an increase in population density would increase the spread of the coronavirus. This was an intuitive result since a higher population density may increase one's exposure to others, increasing the risk of spread.

PD also appeared in interaction terms with HAQI and HYGIENE. Both terms were significant at a 1% level of significance. The interaction of PD and HYGIENE had a positive estimate, while the interaction of PD and HAQI had a negative estimate. It implied that with better and more accessible healthcare, the spread of the coronavirus was likely to reduce despite a high population density. In contrast, increased access to basic hygiene services but high population density could lead to increased spread of the coronavirus.

From the results of equation two, one could observe that HAQI, $HAQI^2$, $HAQI^3$, SANITATION, $SANITATION^2$, and $SANITATION^3$ were all significant at a 5% level of significance. Moreover, the estimates of SANITATION, $SANITATION^3$, HAQI, and $HAQI^3$ were positive, while the estimates of $HAQI^2$ and $SANITATION^2$ were negative. Let us first understand the case of SANITATION by observing its transformation plot. Figure 2 shows the plot of SL versus the cubic transformation of SANITATION.

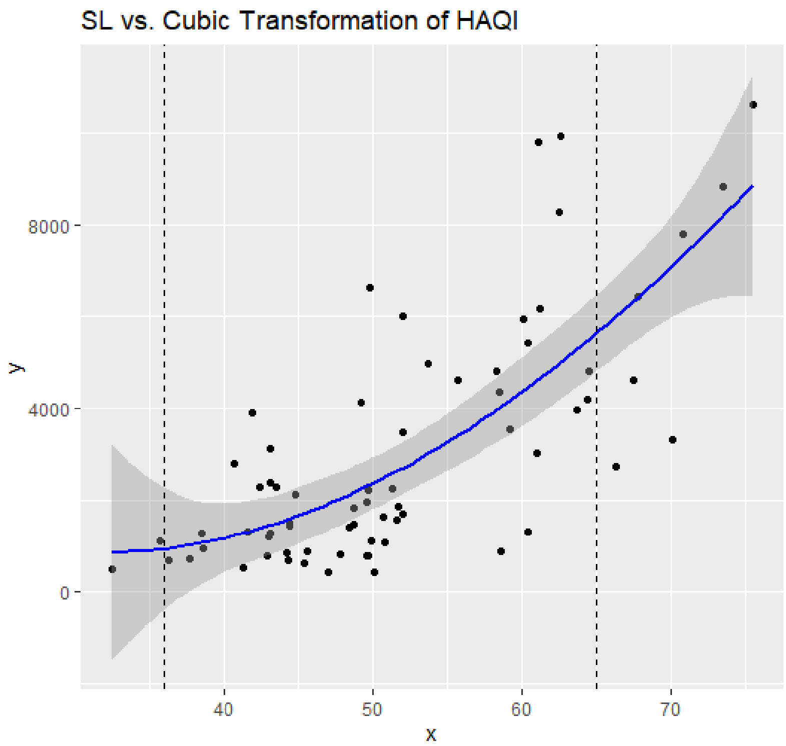
Figure 2: SL vs. Cubic Transformation of SANITATION



As one can observe from the plot, if access to sanitation services is increased to about 20% of the population (the first inflection point), health improves and so does the standard of living. However, investing in a further increase of sanitation services might be costly, which might reduce the short-term standard of living. Therefore, when 20% to 80% of the population has access to basic sanitation, the increase in the long-term standard of living is very slow. This explains the small but negative estimate of $SANITATION^2$. Nonetheless, an even more increase in the access to sanitation services will improve overall health and wellbeing, significantly increasing a country's standard of living.

Similarly, let us understand the case of HAQI by observing its transformation plot. Figure 3 shows the plot of SL versus the cubic transformation of HAQI.

Figure 3: SL vs. Cubic Transformation of HAQI



Similar to the previous case, increasing the healthcare access and quality index until about 36% (the first inflection point) increases a country's standard of living. However, investing in a better and more accessible healthcare system might be costly, hence decreasing its short-term standard of living. Therefore, when HAQI is between 36% to 85%, the increase in the long-term standard of living is slow, as indicated by the small and negative estimate of $HAQI^2$. Nonetheless, an even better and accessible healthcare system would significantly increase a country's standard of living.

Moreover, the interaction term of HYGIENE and SANITATION was significant at a 1% level of significance. The small but negative estimate of the term suggested that the combined increment in the access to basic hygiene and sanitation services was associated with a slow decrease in a country's standard of living. Again, this could be associated with the cost of investment.

From these results, one can claim that economic policies aiming to reduce the spread of Covid-19 must concentrate on increasing access to hygiene services, improving the quality and accessibility of healthcare, and reducing population density. Similarly, economic policies targeting to improve a country's standard of living must increase access to sanitation services, improve the quality and accessibility of healthcare, and limit the combined spending on hygiene and sanitation services.

5. Conclusion

This paper analyzed the significance of the hygiene service industry on economic development and the spread of the coronavirus. Specifically, access to basic hygiene, sanitation, and safe water for drinking were chosen as factors of the hygiene service industry. To predict the spread of Covid-19, the average number of cases was chosen as a response, while to predict economic development, the standard of living was chosen.

Econometric analysis was performed using cross-county data of 70 countries. The data was collected from sources such as *The World Bank* and *UNICEF*. The data collected supported my hypotheses in claiming that access to basic sanitation significantly predicts economic development, access to basic hygiene significantly predicts the spread of the virus, while access to safe water does not predict any of those.

From these results, several policy implications were found. It was found that to reduce the spread of Covid-19, economic policies must focus on increasing access to hygiene services, improving the quality and accessibility of healthcare, and reducing population density. Moreover, to improve a country's standard of living, policies must attempt to increase access to sanitation services, improve the quality and accessibility of healthcare, and limit the combined spending on hygiene and sanitation services.

Given that this was a preliminary study, this research had four main limitations: Firstly, out of 195 countries in total, data for only 70 could be found. Hence, the sample size was relatively small. Secondly, the data used observations from different years. Thirdly, this research did not attempt to remove outliers due to its already small sample size. Lastly, IV regression involved the use of weak instruments. However, instrumental variable regression helped me overcome endogeneity bias, which made this research reliable.

References

American Cleaning Institute. (2020). *Rising to an Unprecedented Challenge: The cleaning products industry's response to COVID-19*. https://www.cleaninginstitute.org/sites/default/files/documents/Coronavirus/2020ACIImpactReport_COVID19.pdf

American Society for Parenteral and Enteral Nutrition. (2020). *https://www.nutritioncare.org/uploadedFiles/Documents/Guidelines_and_Clinical_Resources/COVID19/COVID19%20Patient%20Nutrition%20Paper.pdf*. ASPEN. https://www.nutritioncare.org/uploadedFiles/Documents/Guidelines_and_Clinical_Resources/COVID19/COVID19%20Patient%20Nutrition%20Paper.pdf

Appel, C., Beltekian, D., Gavrilov, D., Giattino, C., Hasell, J., Macdonald, B., Mathieu, E., Ortiz-Ospina, E., Ritchie, H., & Roser, M. (2021). *Covid-19-data*. GitHub. https://github.com/owid/covid-19-data/tree/master/public/data

Butler, G., Pilotto, R. G., Hong, Y., & Mutambatsere, E. (2020). *The Impact of COVID-19 on the Water and Sanitation Sector*. International Finance Corporation. https://www.ifc.org/wps/wcm/connect/126b1a18-23d9-46f3-beb7-047c20885bf6/The+Impact+of+COVID_Water%26Sanitation_final_web.pdf?MOD=AJPERES&CVID=ncaG-hA

Donde, O. O., Atoni, E., Muia, A. W., & Yillia, P. T. (2021). COVID-19 pandemic: Water, sanitation and hygiene (WASH) as a critical control measure remains a major challenge in low-income countries. *Water Research*, *191*, 116793. https://doi.org/10.1016/j.watres.2020.116793

EnsembleIQ. (2020, April). Store Brands Magazine. *Issuu*, *0420*. https://issuu.com/ensembleiq/docs/sbr-0420?fr=sZmIzYTgwODQ2MQ

Euromonitor International. (2020, April). *The Impact of Coronavirus on Tissue and Disposable Hygiene*. https://www.euromonitor.com/the-impact-of-coronavirus-on-tissue-and-disposable-hygiene/report

Glen, S. (2020, December 28). *Instrumental Variable: Definition & Overview*. Statistics How To. https://www.statisticshowto.com/instrumental-variable/

Global Change Data Lab. (2016). *Healthcare Access and Quality Index*. Our World in Data. https://ourworldindata.org/grapher/healthcare-access-and-quality-index

Global Industry Analysts Inc. (2020, September). *Hand Hygiene - Global Market Trajectory & Analytics* (No. 5140971). Research and Markets. https://www.researchandmarkets.com/reports/5140971/hand-hygiene-global-market-trajectory-and?utm_source=BW&utm_medium=PressRelease&utm_code=27plml&utm_campaign=1487694+-+%241.2+Billion+Worldwide+Hand+Hygiene+Industry+to+2027+-+Impact+of+COVID-19+on+the+Market&utm_exec=jamu273prd

Hygiene Matters. (2017). *Hygiene - a catalyst for economic growth*. Essity (publ). https://reports.essity.com/2016-17/hygiene-matters-report/the-value-of-hygiene/hygiene-a-catalyst-for-economic-growth.html

International Finance Corporation. (2021). *IFC - International Finance Corporation*. IFC. https://www.ifc.org/wps/wcm/connect/corp_ext_content/ifc_external_corporate_site/home

Keller, S. (2019). *Water, Sanitation and Economy*. SSWM - Find Tools for Sustainable Sanitation and Water Management! https://sswm.info/arctic-wash/module-1-introduction/further-resources-sustainability-relation-water-sanitation/water%2C-sanitation-and-economy

Kemeny, T. (2007, September). *Sanitation and economic development – Making an economic case for the MD*. WaterAid America. https://washmatters.wateraid.org/publications/sanitation-and-economic-development-2007

Kraay, A. (2019). The World Bank Human Capital Index: A Guide. *The World Bank Research Observer*, *34*(1), 1–33. https://doi.org/10.1093/wbro/lkz001

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, *107*(2), 407–437. https://doi.org/10.2307/2118477

Moore, L. D., Robbins, G., Quinn, J., & Arbogast, J. W. (2021). The impact of COVID-19 pandemic on hand hygiene performance in hospitals. *American Journal of Infection Control*, *49*(1), 30–33. https://doi.org/10.1016/j.ajic.2020.08.021

Research and Markets Ltd. (2020, March). *Impact of COVID 19 on the Hygiene Industry - Research and Markets*. Research and Markets Ltd 2021. https://www.researchandmarkets.com/reports/5013569/impact-of-covid-19-on-the-hygiene-industry?utm_source=dynamic&utm_medium=BW&utm_code=jh8kh4&utm_campaign=1385212+-+Impact+of+COVID-

19+on+the+Hygiene+Industry%2c+2020%3a+Unprecedented+Growth+Spurred+by+Stringent+Hygienic+Guidelines&utm_exec=joca220bwd

Ritchie, H., & Roser, M. (2019, August 25). *Meat and Dairy Production*. Our World in Data. https://ourworldindata.org/meat-production

Sakia, R. M. (1992). The Box-Cox Transformation Technique: A Review. *The Statistician*, *41*(2), 169. https://doi.org/10.2307/2348250

Statista. (2020, November 26). *U.S. online sales growth of health and hygiene products March 2020*. https://www.statista.com/statistics/1112767/impact-of-coronavirus-on-online-sales-values-of-health-and-hygiene-products-in-the-us/

Stock, J. H., & Watson, M. W. (2008). *Introduction to Econometrics* (3rd ed.). Pearson/Addison Wesley.

The World Bank. (2018). *GDP per capita (current US$) | Data*. World Bank Data. https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

The World Bank. (2021a). *Glossary | DataBank*. The World Bank Group. https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NY.GDP.PCAP.CD

The World Bank. (2021b). *Human Capital*. World Bank Database. https://www.worldbank.org/en/publication/human-capital

Tripathi, A. (2019, June 16). *What is stepAIC in R? - Ashutosh Tripathi*. Medium. https://ashutoshtr.medium.com/what-is-stepaic-in-r-a65b71c9eeba#:%7E:text=AIC%20stands%20for%20Akaike%20Information%20Crit eria.&text=Hence%20we%20can%20say%20that,does%20not%20have%20any%20significance.

United Nations International Children’s Emergency Fund. (2016). *Hygiene and Hand Washing Statistics*. UNICEF DATA. https://data.unicef.org/topic/water-and-sanitation/hygiene/

Van Minh, H., & Hung, N. V. (2011). Economic Aspects of Sanitation in Developing Countries. *Environmental Health Insights*, *5*, EHI.S8199. https://doi.org/10.4137/ehi.s8199

Xinhua. (2020, February 7). *Rumor Buster: Can vegetarian diet prevent COVID-19 infections? - Xinhua | English.news.cn*. Xinhua Net. http://www.xinhuanet.com/english/2020-07/02/c_139182885.htm

Appendix

A. Data Documentation

SL	Variable Name	Definition/Description	Unit	Type / Range
1	CASES	The average number of new coronavirus positive cases for a country, on a seven-day rolling basis as of March 1, 2021.	Number of cases	Long Decimal Numbers
2	HCI	*Human Capital Index:* The average levels of education and health of an individual in a country.	N/A	0 to 100
3	MEAT	Meat consumption per capita of a country.	Kilograms	Long Decimal Numbers
4	PD	*Population density:* The number of people per square kilometer.	Number of people	Long Decimal Numbers
5	HYGIENE	The percentage of the population that has access to at least basic hygiene services.	Percentage of people	Percentage
6	SANITATION	The percentage of the population that has access to at least basic sanitation services.	Percentage of people	Percentage
7	WATER	The percentage of the population that has access to safe water for drinking.	Percentage of people	Percentage
8	SL	*Standard of living:* The GDP per capita that is PPP adjusted, for current U.S. dollars.	Current U.S. Dollars	Long Decimal Numbers
9	HAQI	*Health Access and Quality Index:* The measure of the quality and accessibility of healthcare institutions.	N/A	0 to 100

B. Transformations with response CASES

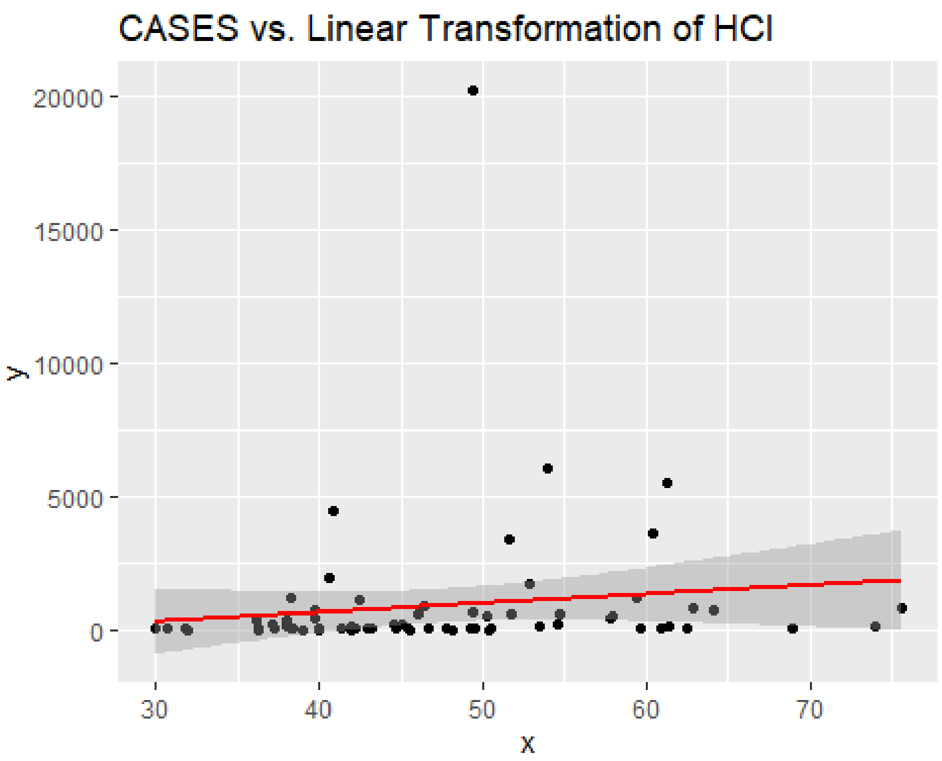
```
y = CASES

# -----------------------------------------------------------------------
x = HCI

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))
```

```
## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0588 0.8092
```

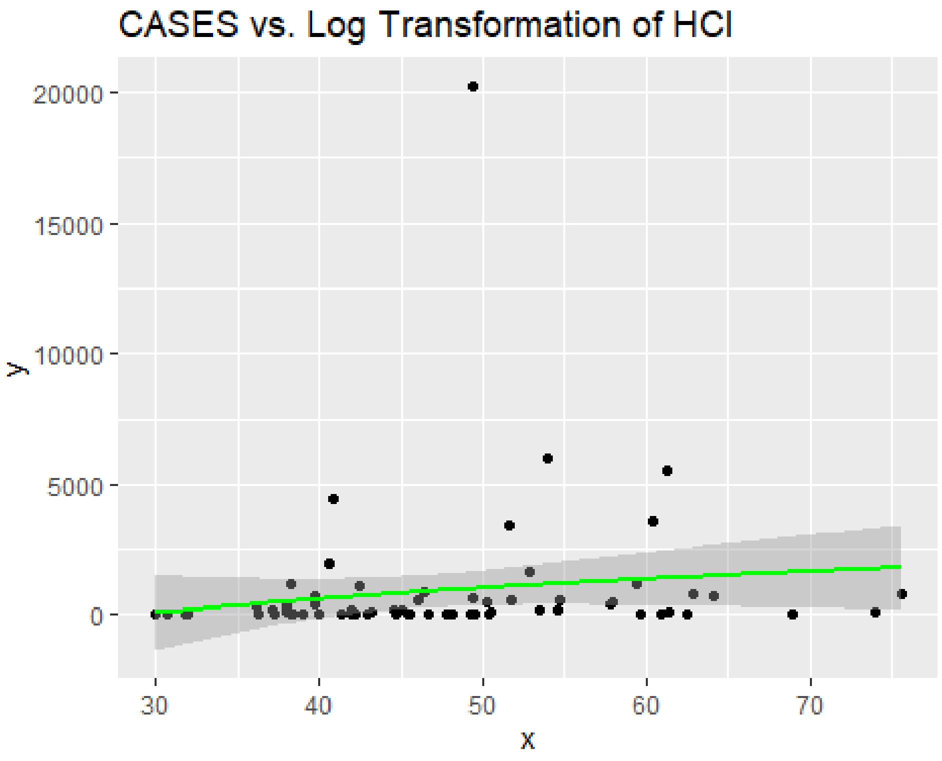
```
# Since p-value = 0.8092 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))
```

```
## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     66
## 2     68 -2 1.6274 0.2042
```

```
# Since p-value = 0.2042 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
CASES vs. Linear Transformation of HCI")
```

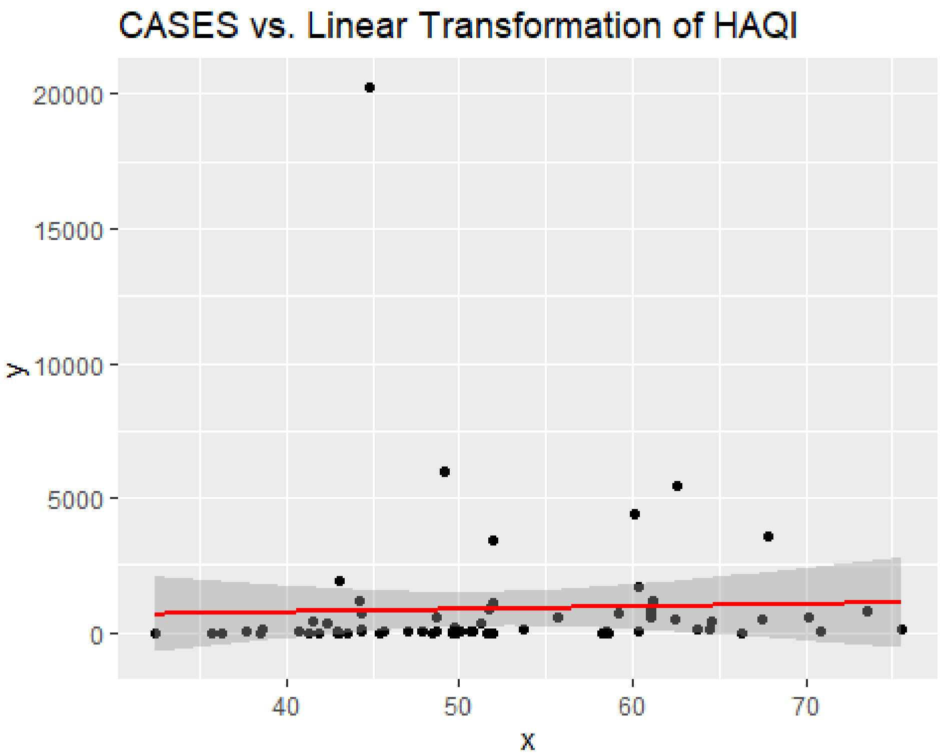



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of HCI")
```

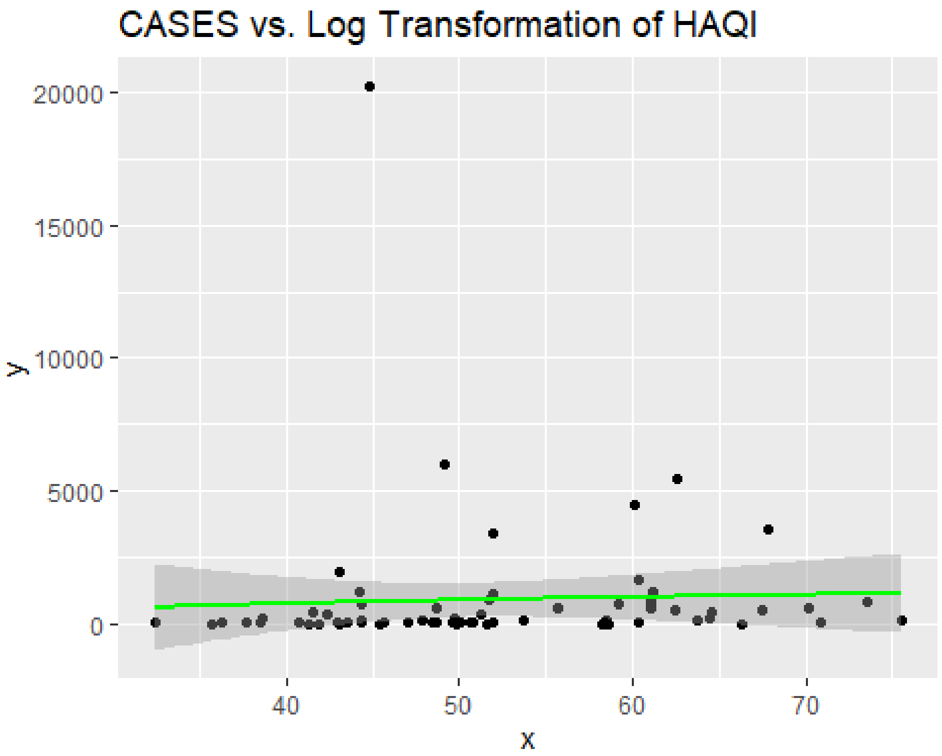



```
# From graphs, I can conclude that the linear transformation is the best
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
x = HAQI

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0038  0.951

# Since p-value = 0.951 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
```

```
## 1     66
## 2     68 -2 1.5264 0.2249

# Since p-value = 0.2249 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
CASES vs. Linear Transformation of HAQI")
```

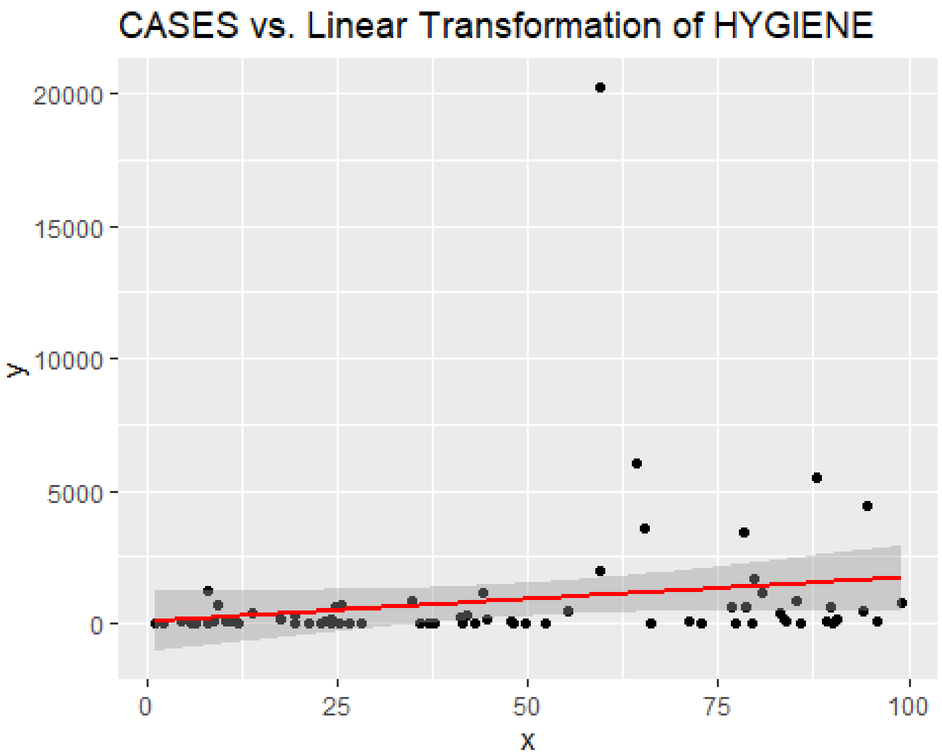



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of HAQI")
```

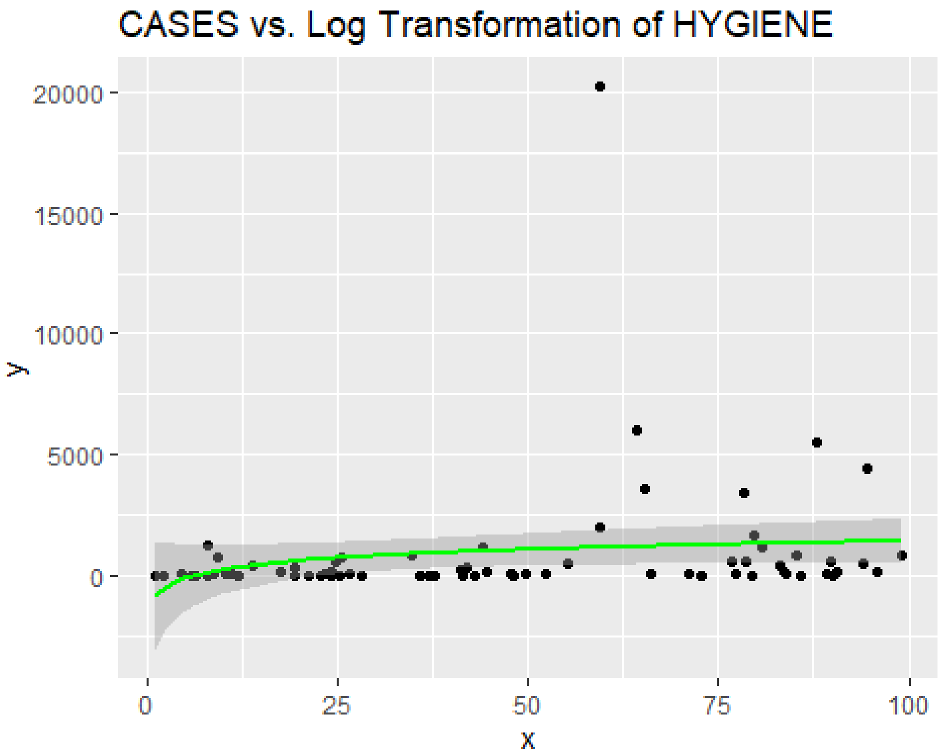



```
# From graphs, I can conclude that the linear transformation is the best
# -----------------------------------------------------------------------

# -----------------------------------------------------------------------
x = HYGIENE

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0423 0.8377

# Since p-value = 0.8377 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
```

```
## 1     66
## 2     68 -2 1.2669 0.2885

# Since p-value = 0.2885 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
CASES vs. Linear Transformation of HYGIENE")
```

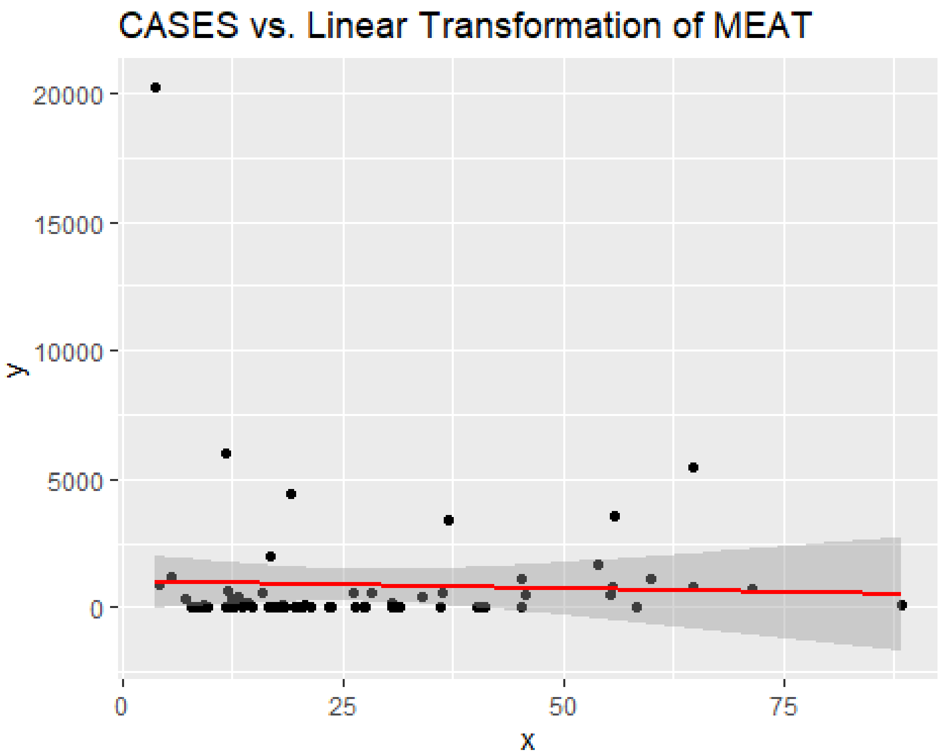



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of HYGIENE")
```

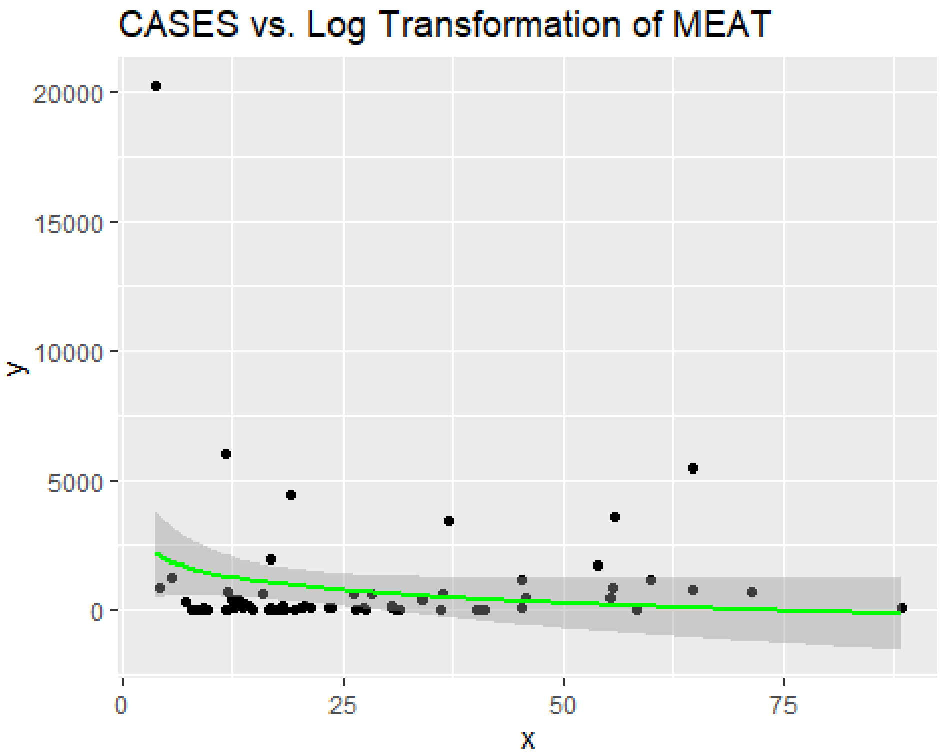



```
# From graphs, I can conclude that the linear transformation is the best
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
x = MEAT

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0288 0.8658

# Since p-value = 0.8658 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
```

```
## 1     66
## 2     68 -2 1.7108 0.1886

# Since p-value = 0.1886 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
CASES vs. Linear Transformation of MEAT")
```

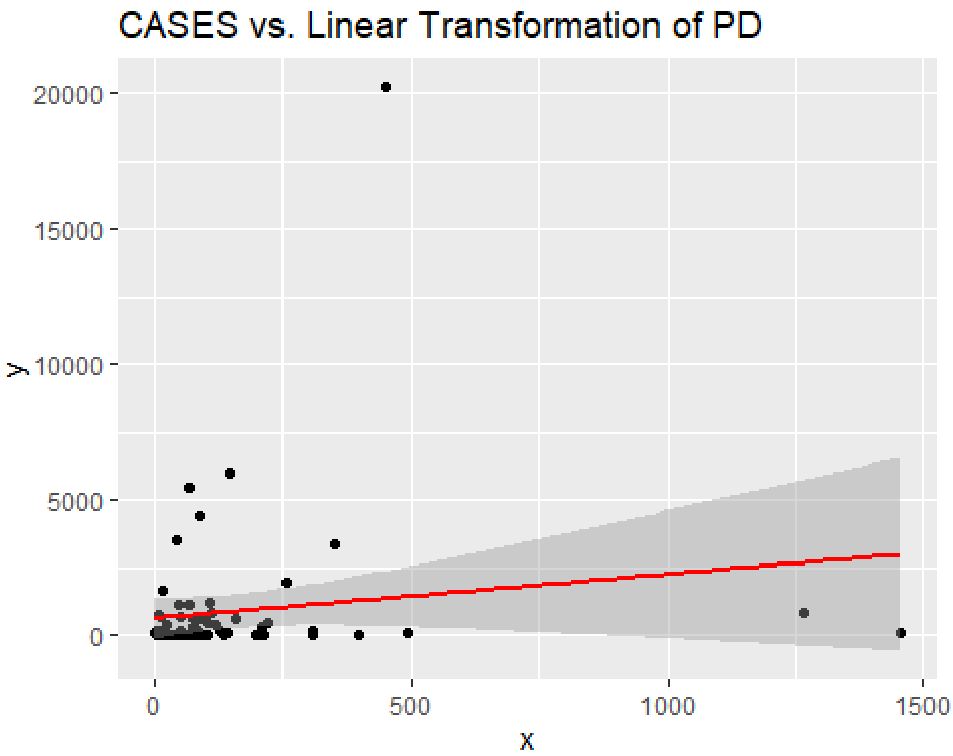



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of MEAT")
```

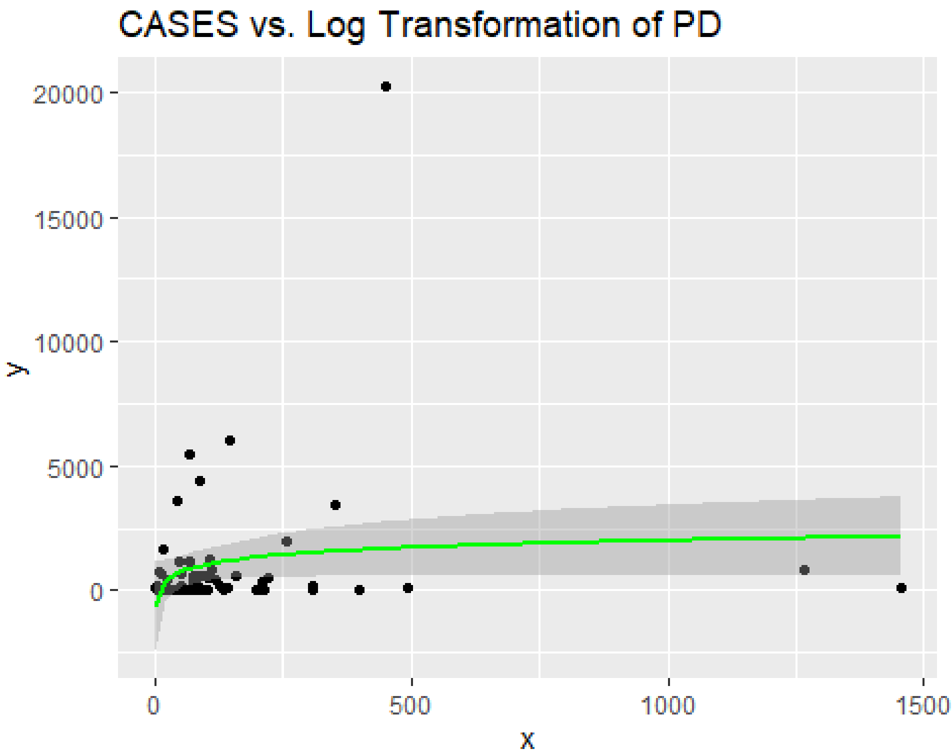



```r
# From graphs, I can conclude that the linear transformation is the best
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
x = PD

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0823 0.7751

# Since p-value = 0.7751 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
```

```
## 1     66
## 2     68 -2 0.8107 0.4489

# Since p-value = 0.4489 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
CASES vs. Linear Transformation of PD")
```



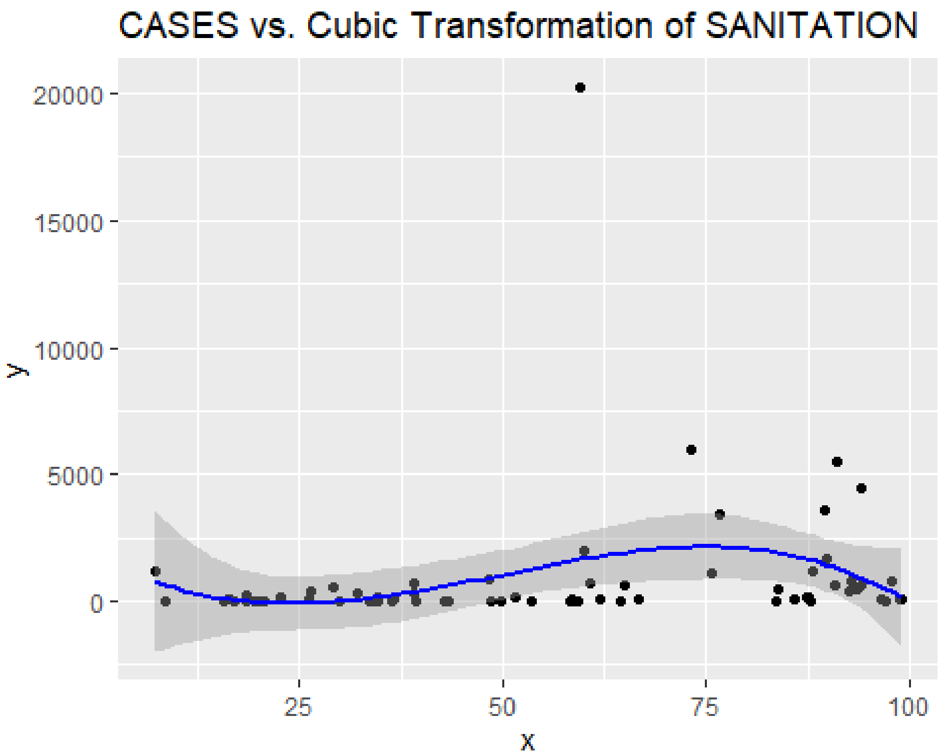

```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of PD")
```



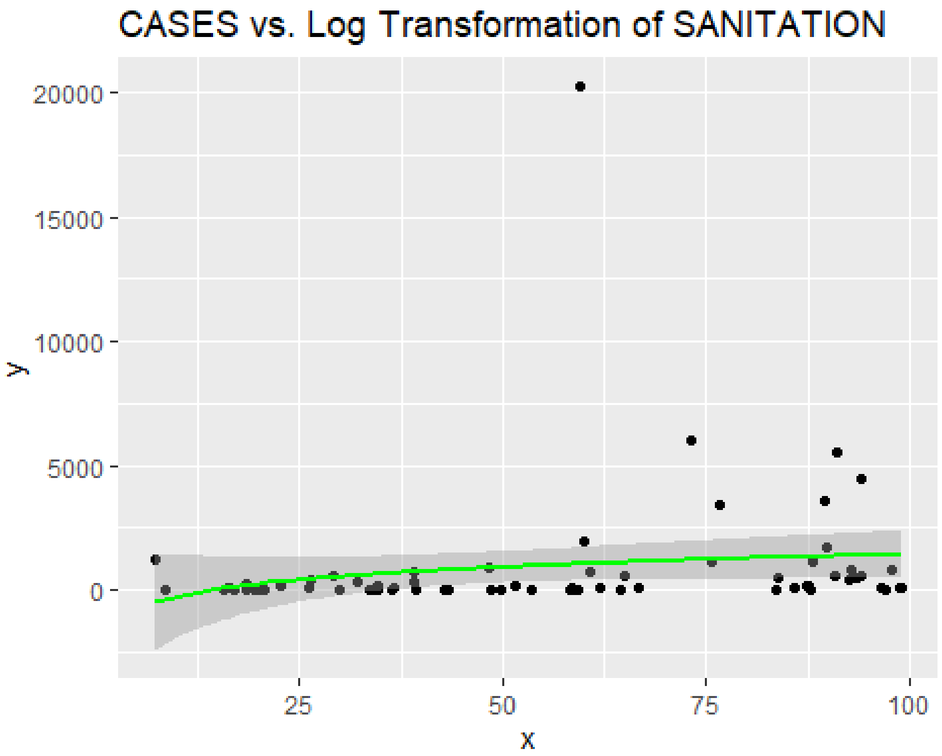

```
# From graphs, I can conclude that the linear transformation is the best
# -----------------------------------------------------------------------

# -----------------------------------------------------------------------
x = SANITATION

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0451 0.8325

# Since p-value = 0.8325 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F  Pr(>F)
## 1     66
```

```
## 2     68 -2 3.8924 0.02524 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Since p-value = 0.02524 < 0.05, Cubic model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2) + I(x^3), color =
"blue") + ggtitle("CASES vs. Cubic Transformation of SANITATION")
```

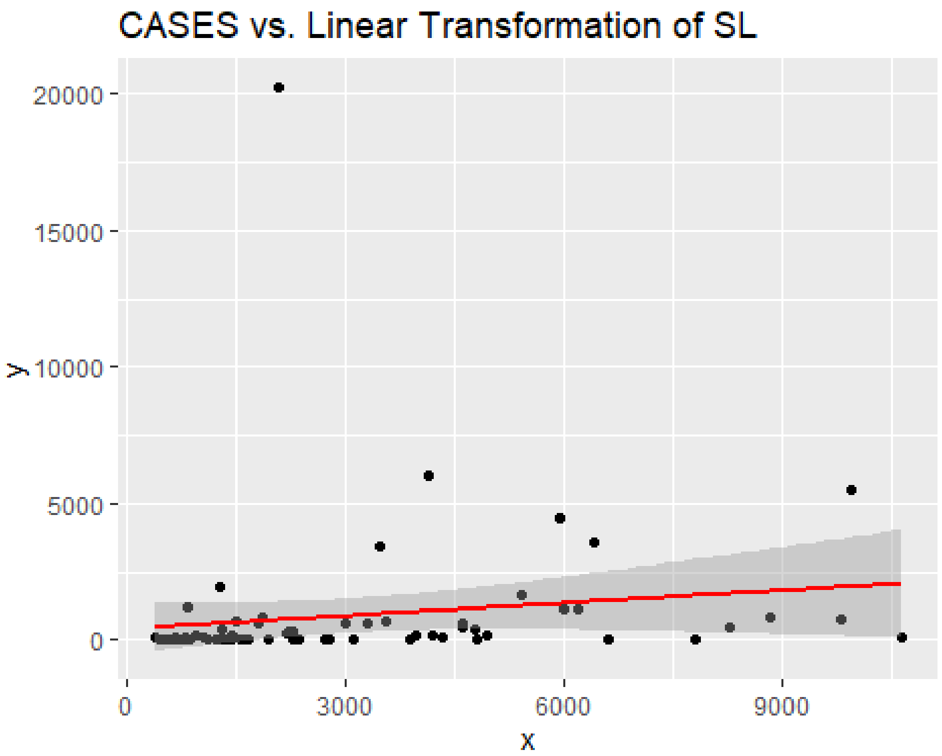



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of SANITATION")
```

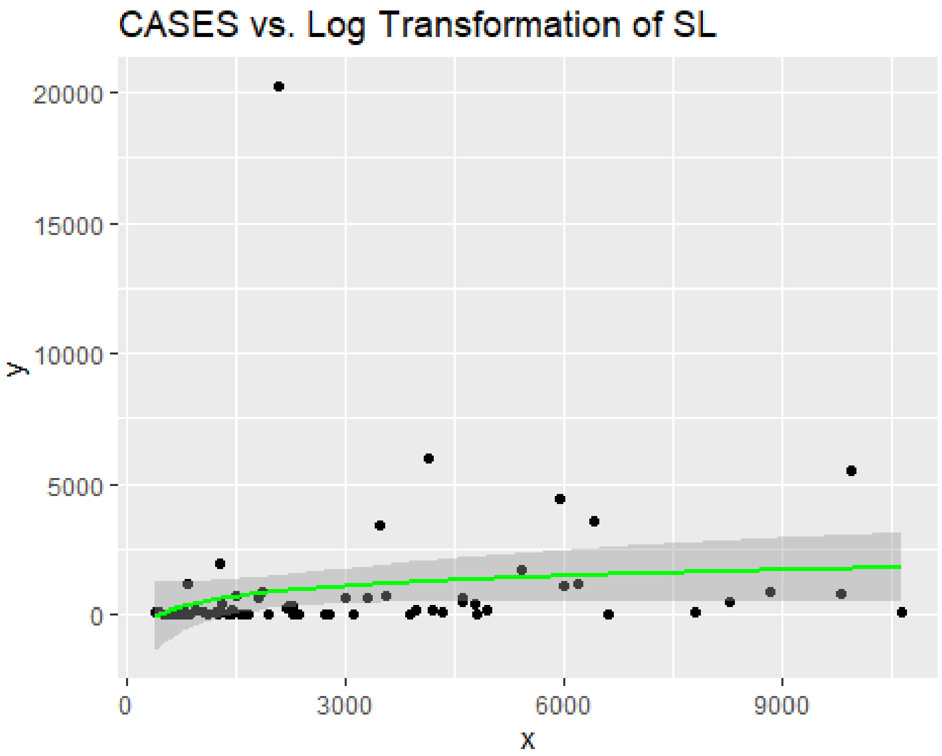



```
# From graphs, I can conclude that the cubic transformation is the best
# -----------------------------------------------------------------------

# -----------------------------------------------------------------------
x = SL

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0111 0.9164

# Since p-value = 0.9164 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
```

```
## 1     66
## 2     68 -2 0.6045 0.5494

# Since p-value = 0.5494 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
CASES vs. Linear Transformation of SL")
```



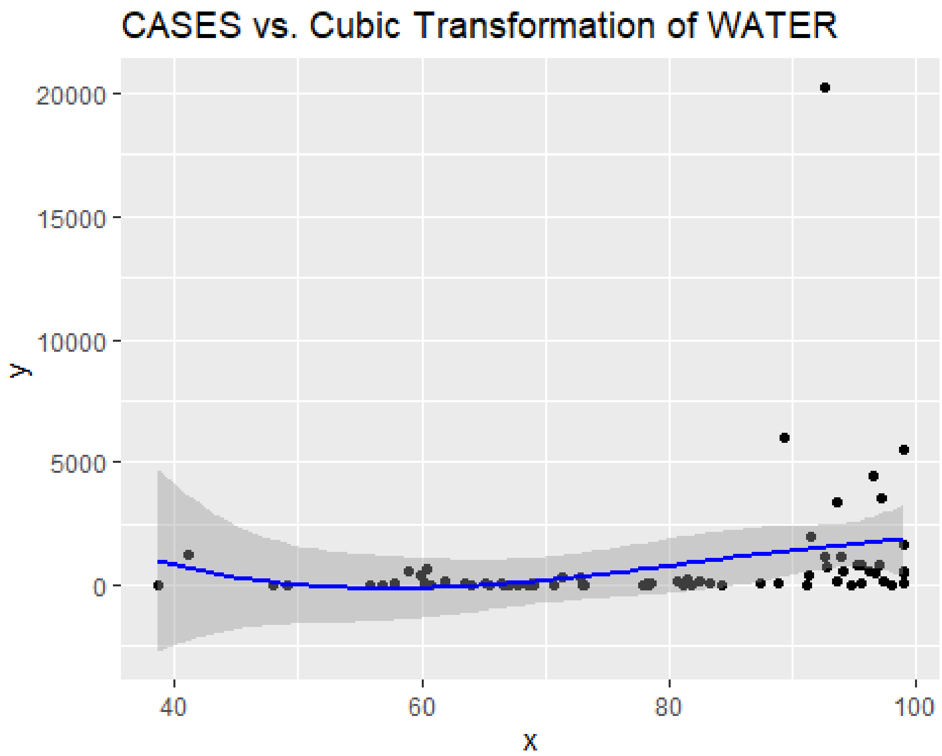

```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of SL")
```



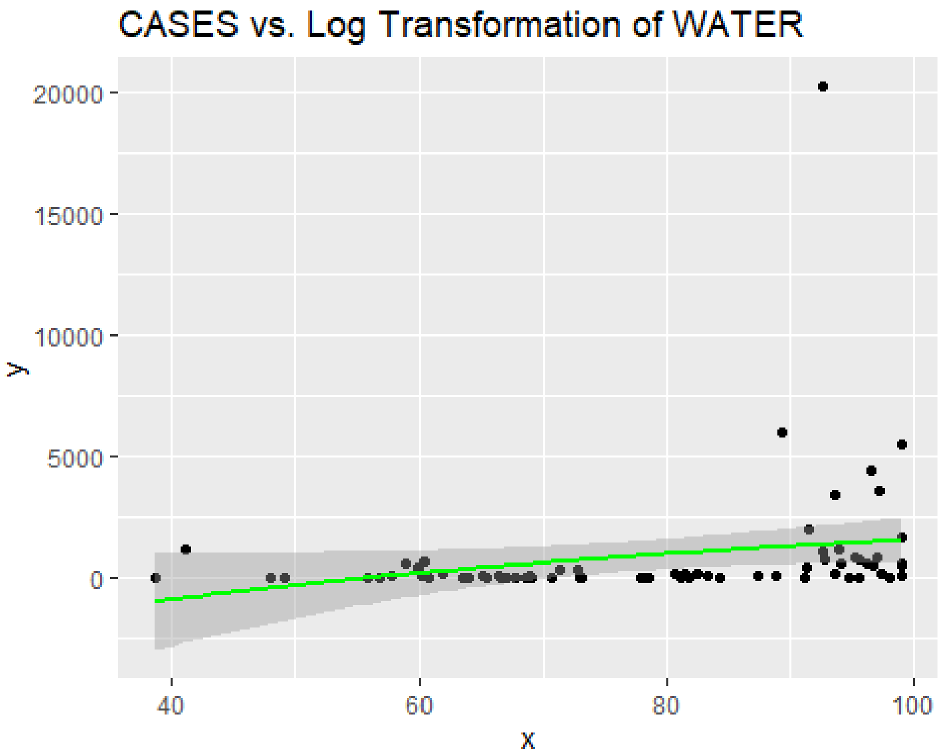

```
# From graphs, I can conclude that the log transformation is the best
# ------------------------------------------------------------------------

# ------------------------------------------------------------------------
x = WATER

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0103 0.9194

# Since p-value = 0.9194 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F  Pr(>F)
## 1     66
```

```
## 2     68 -2 3.4758 0.03671 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Since p-value = 0.03671 < 0.05, Cubic model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2) + I(x^3), color =
"blue") + ggtitle("CASES vs. Cubic Transformation of WATER")
```




```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("CASES vs. Log Transformation of WATER")
```

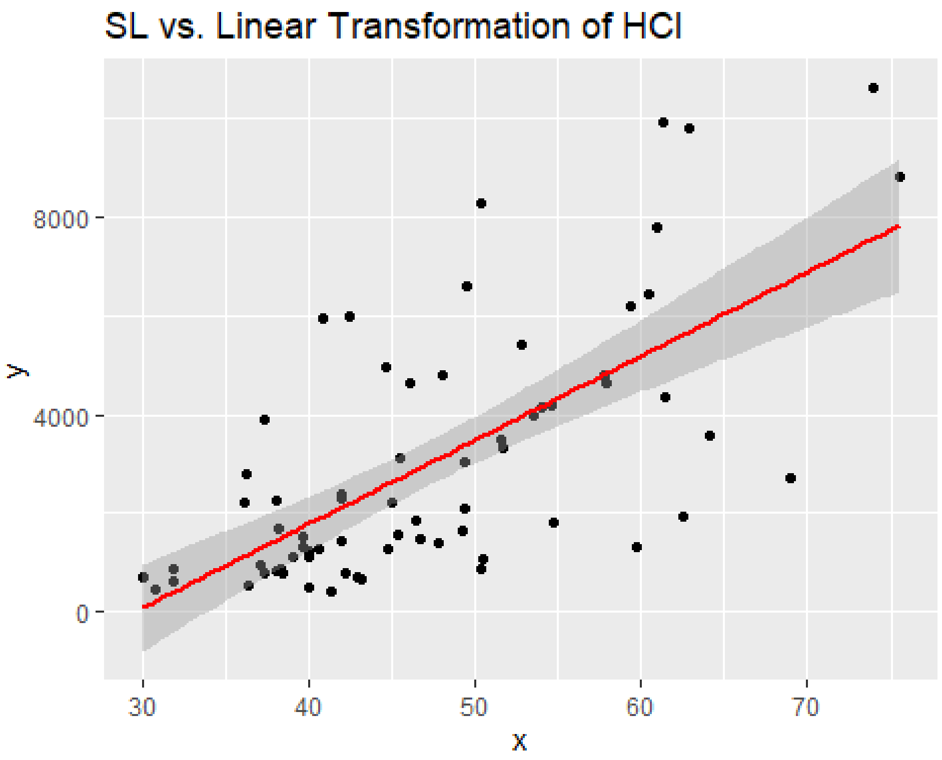



```
# From graphs, I can conclude that the cubic transformation is the best
# ---------------------------------------------------------------------------
```

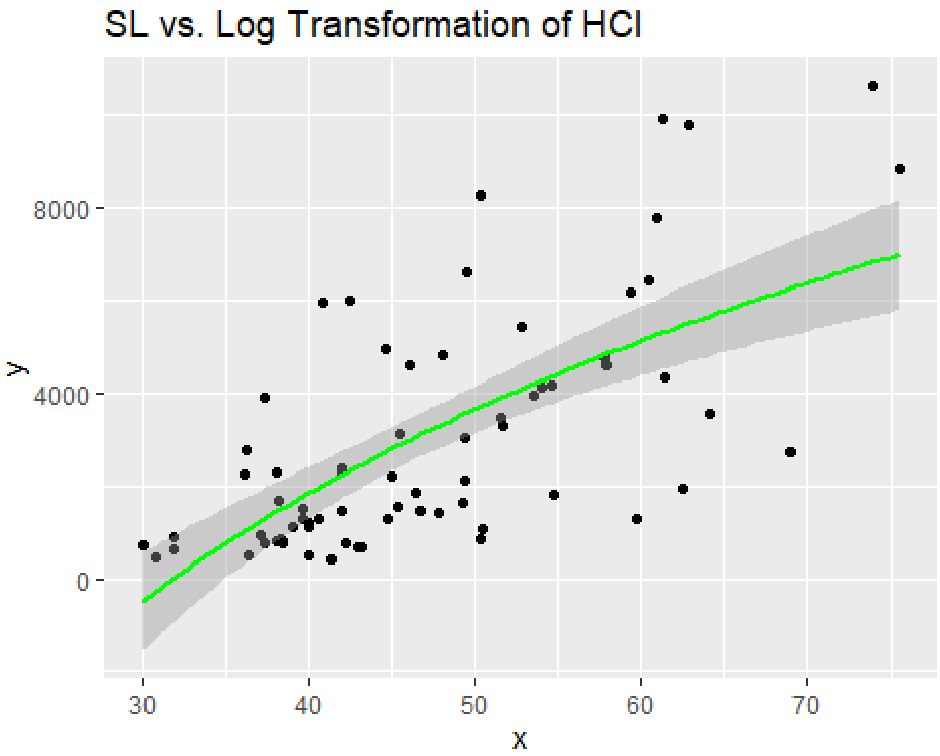
C. Transformations with response SL

```
y = SL

# ---------------------------------------------------------------------------
x = HCI

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0088 0.9257

# Since p-value = 0.9257 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))
```

```
## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     66
## 2     68 -2 0.6141 0.5442

# Since p-value = 0.5442 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
SL vs. Linear Transformation of HCI")
```

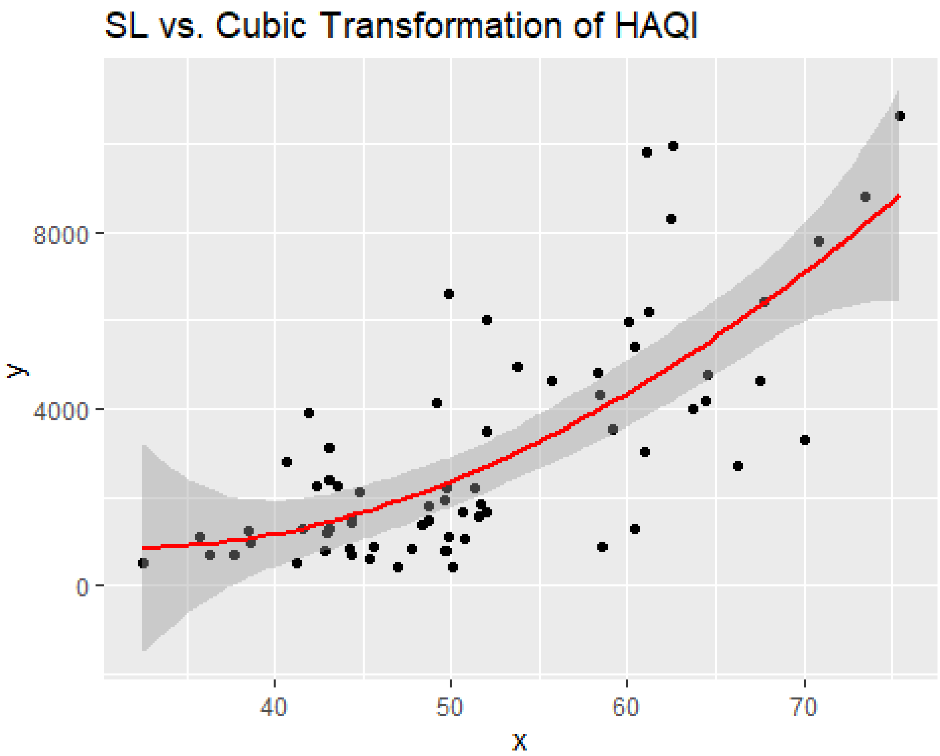



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("SL vs. Log Transformation of HCI")
```

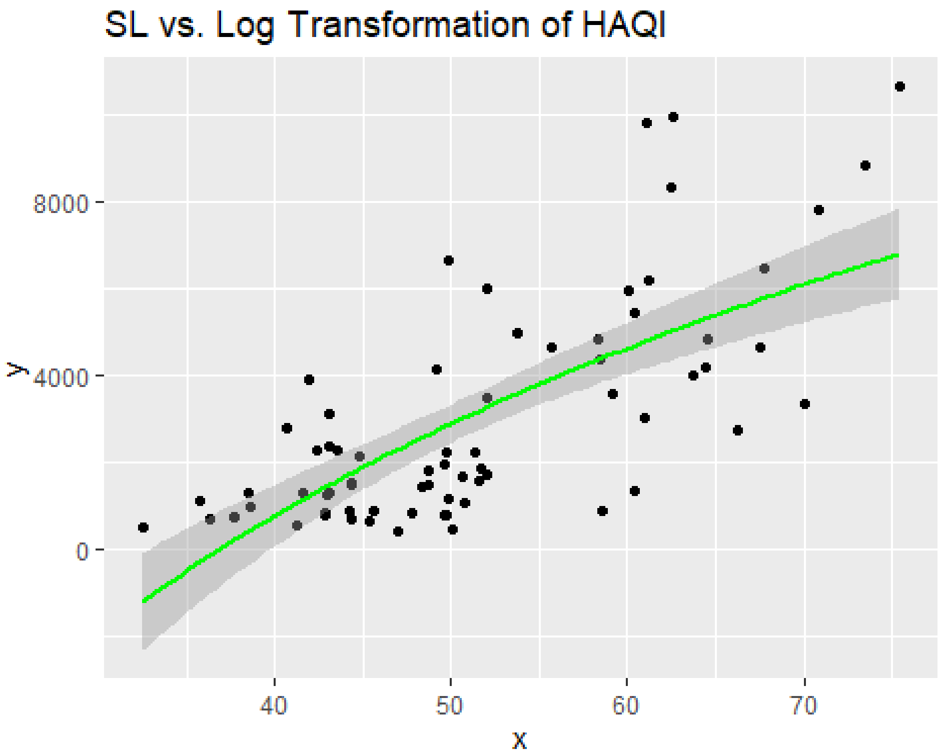



```
# From graphs, I can conclude that the linear transformation is the best
# ------------------------------------------------------------------------

# ------------------------------------------------------------------------
x = HAQI

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0265 0.8712

# Since p-value = 0.8712 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df     F  Pr(>F)
## 1     66
```

```
## 2     68 -2 3.652 0.03131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Since p-value = 0.03131 < 0.05, Cubic model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2) + I(x^3), color =
"red") + ggtitle("SL vs. Cubic Transformation of HAQI")
```

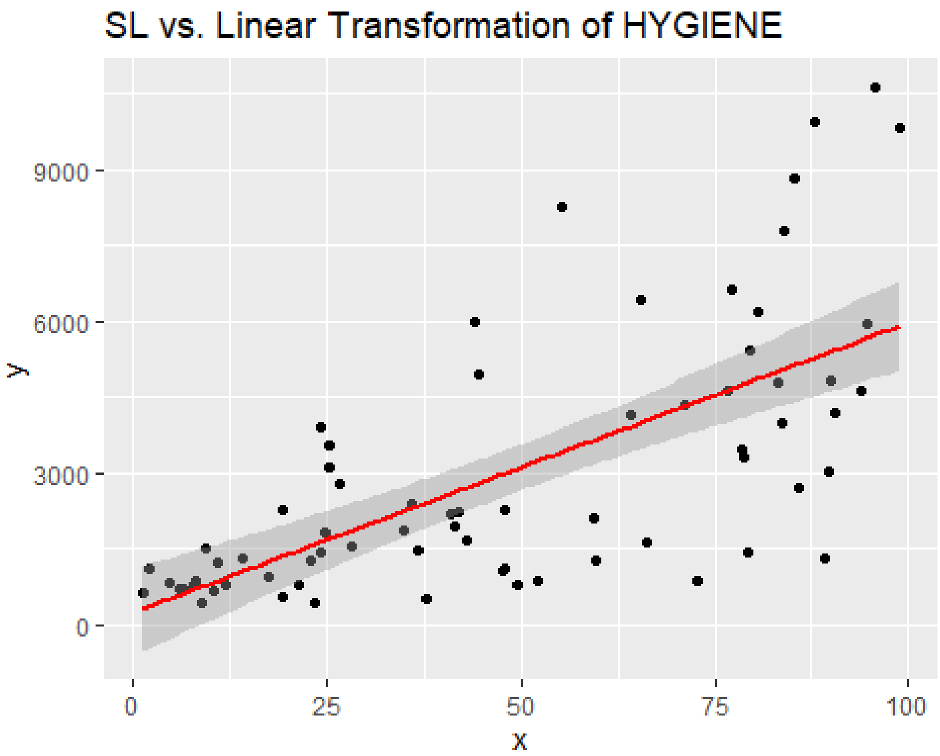



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("SL vs. Log Transformation of HAQI")
```

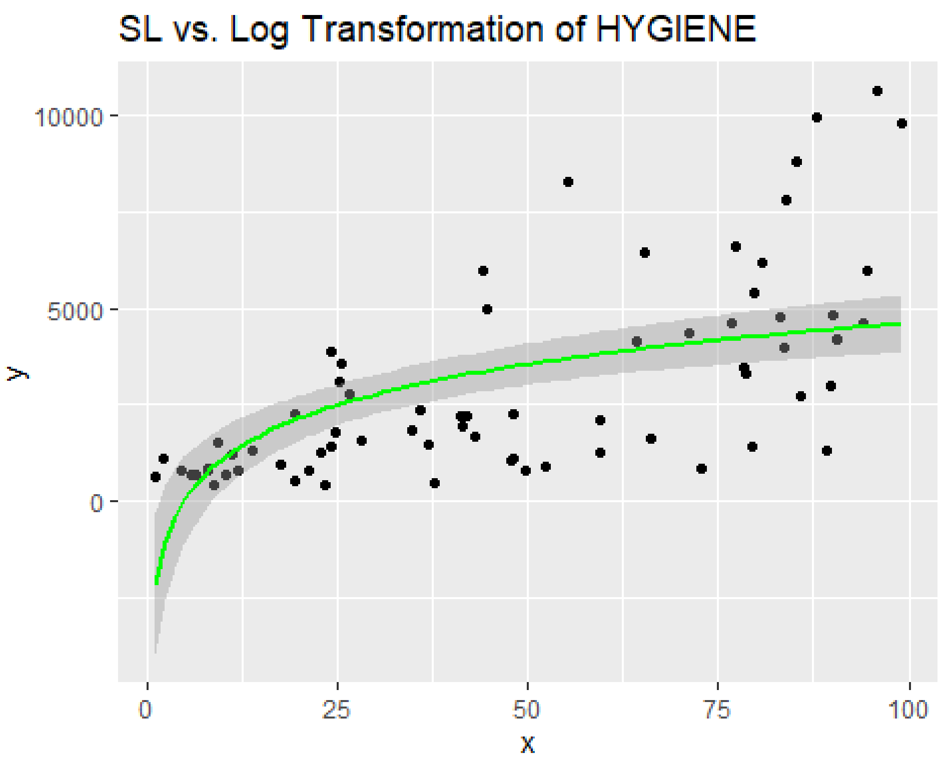



```
# From graphs, I can conclude that the cubic transformation is the best
# ------------------------------------------------------------------------

# ------------------------------------------------------------------------
x = HYGIENE

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0985 0.7547

# Since p-value = 0.7547 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
```

```
## 1     66
## 2     68 -2 1.6334 0.2031

# Since p-value = 0.2031 > 0.05, Linear model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x, color = "red") + ggtitle("
SL vs. Linear Transformation of HYGIENE")
```



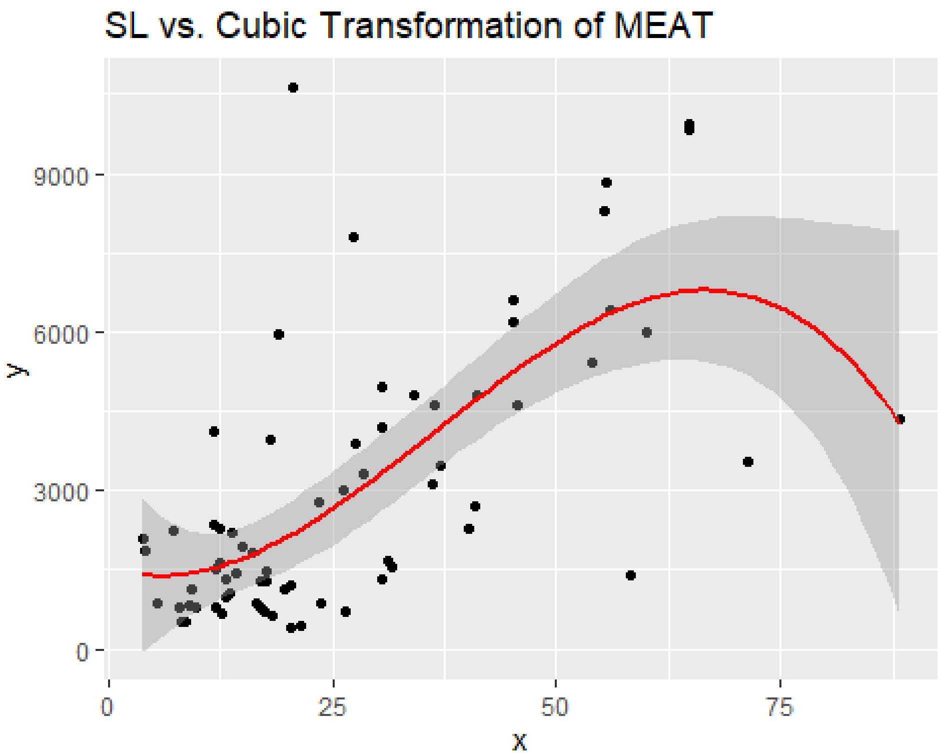

```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("SL vs. Log Transformation of HYGIENE")
```



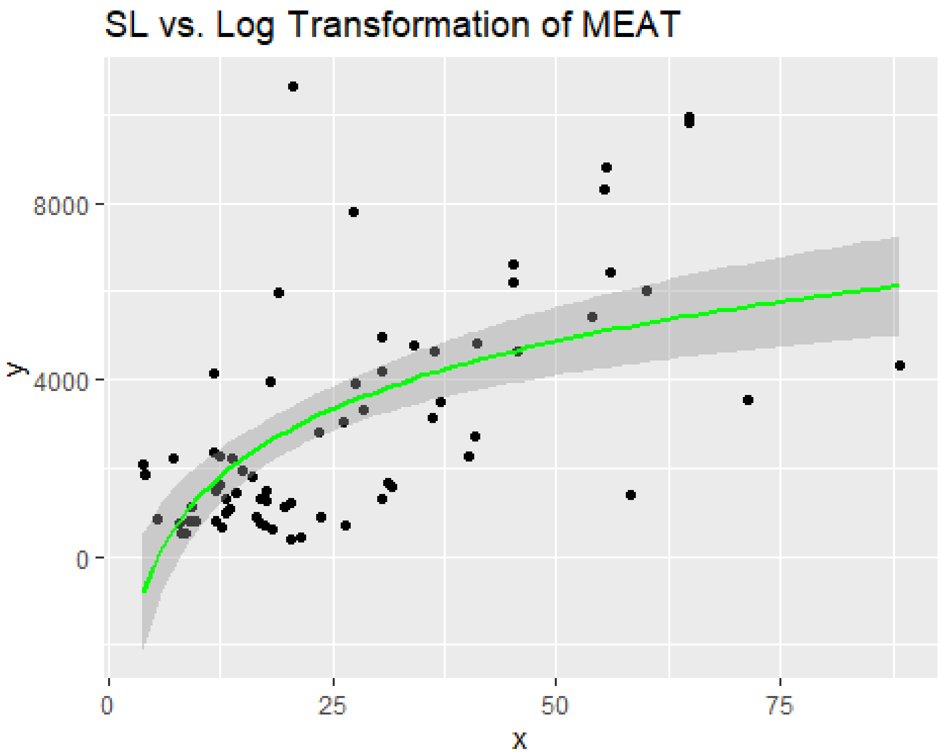

```
# From graphs, I can conclude that the linear transformation is the best
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
x = MEAT

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0863 0.7699

# Since p-value = 0.7699 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F    Pr(>F)
## 1     66
```

```
## 2     68 -2 11.358 5.765e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Since p-value = 5.765e-05 < 0.05, Cubic model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2) + I(x^3), color =
"red") + ggtitle("SL vs. Cubic Transformation of MEAT")
```



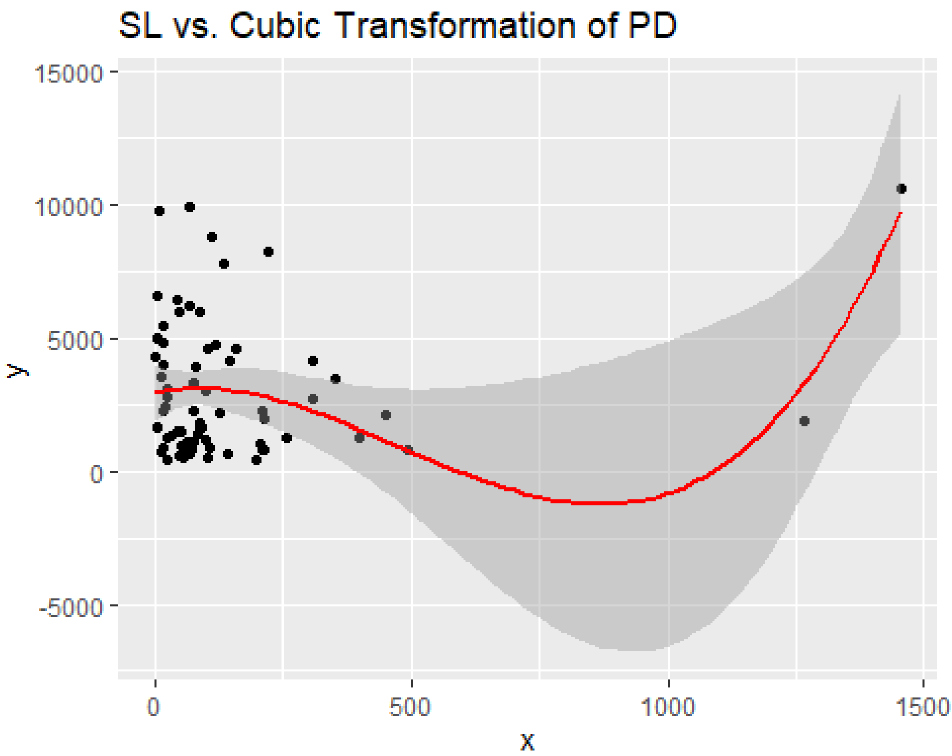

```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("SL vs. Log Transformation of MEAT")
```



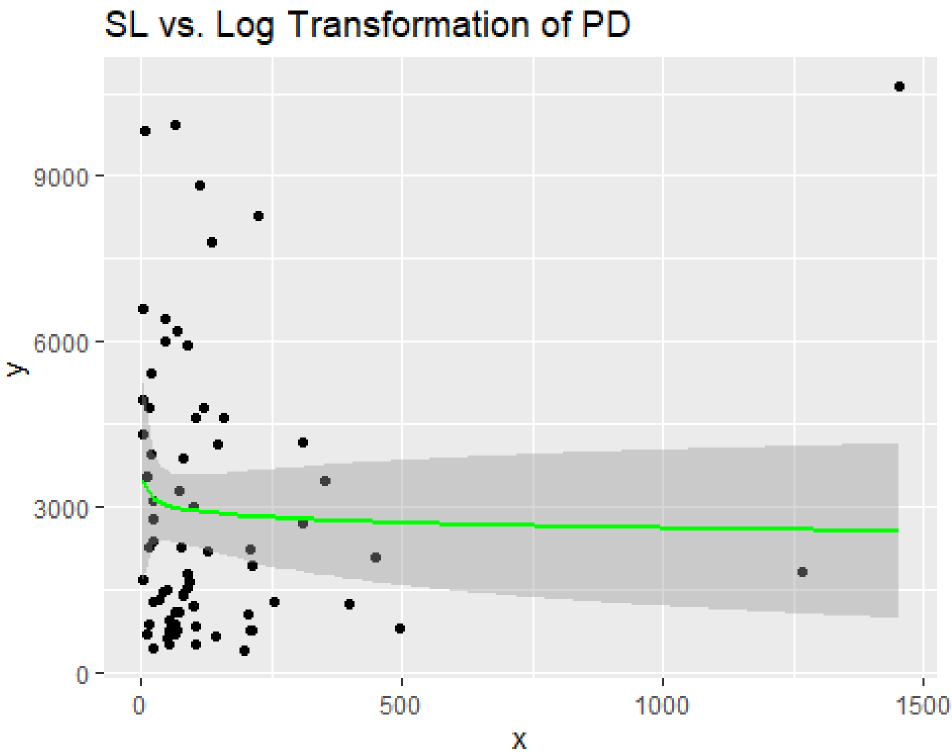

```
# From graphs, I can conclude that the cubic transformation is the best
# ---------------------------------------------------------------------------

# ---------------------------------------------------------------------------
x = PD

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.1725 0.6792

# Since p-value = 0.6792 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F    Pr(>F)
## 1     66
```

```
## 2     68 -2 20.887 9.376e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Since p-value = 9.376e-08 < 0.05, Cubic model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2) + I(x^3), color =
"red") + ggtitle("SL vs. Cubic Transformation of PD")
```

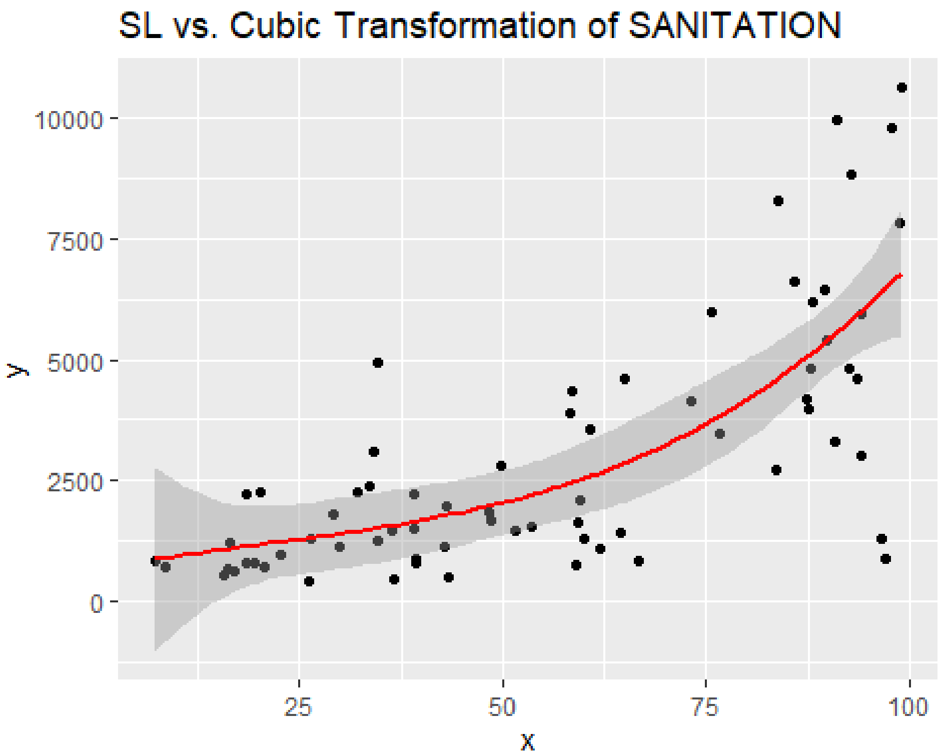



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("SL vs. Log Transformation of PD")
```

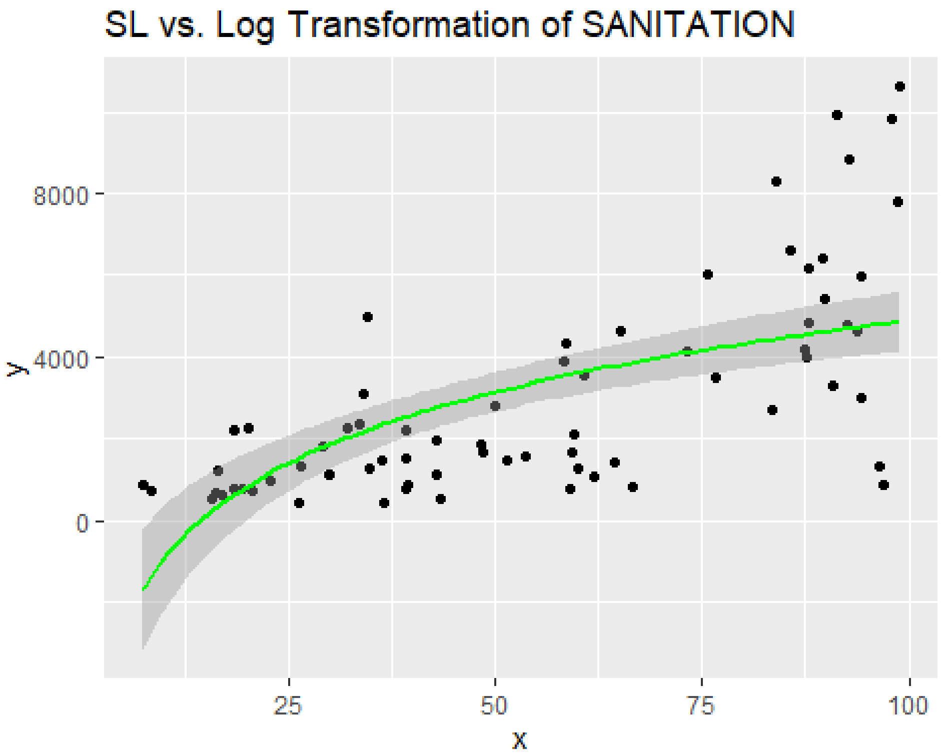



```
# From graphs, I can conclude that the cubic transformation is the best
# ----------------------------------------------------------------------

# ----------------------------------------------------------------------
x = SANITATION

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.1081 0.7434

# Since p-value = 0.7434 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df     F  Pr(>F)
## 1     66
```

```
## 2     68 -2 4.736 0.01197 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Since p-value = 0.01197 < 0.05, Cubic model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2) + I(x^3), color =
"red") + ggtitle("SL vs. Cubic Transformation of SANITATION")
```

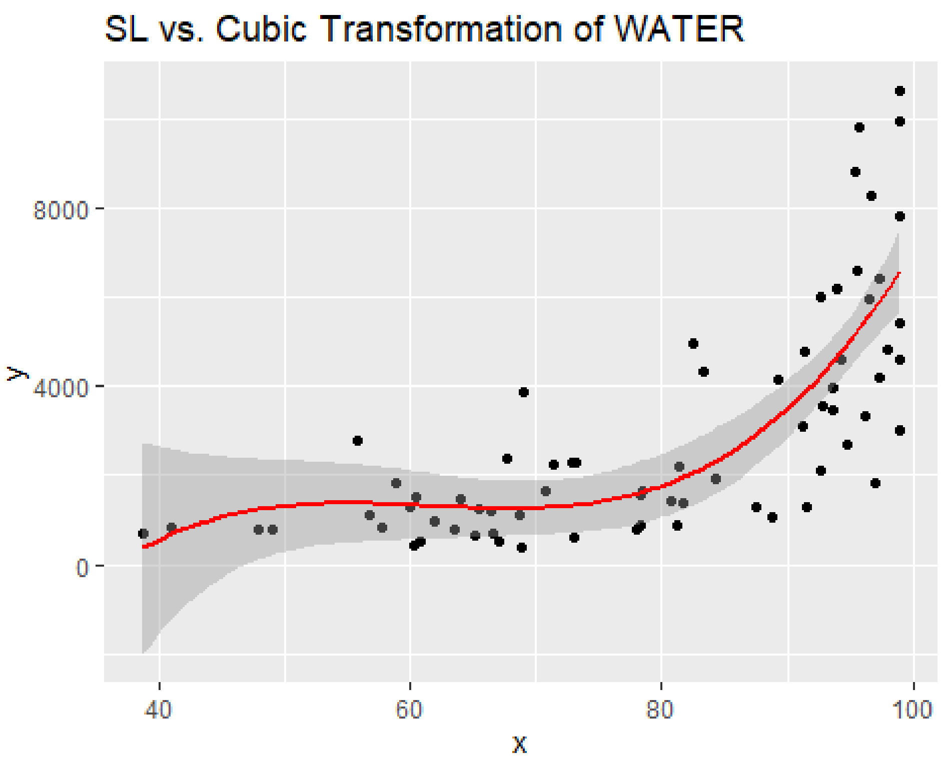



```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("SL vs. Log Transformation of SANITATION")
```

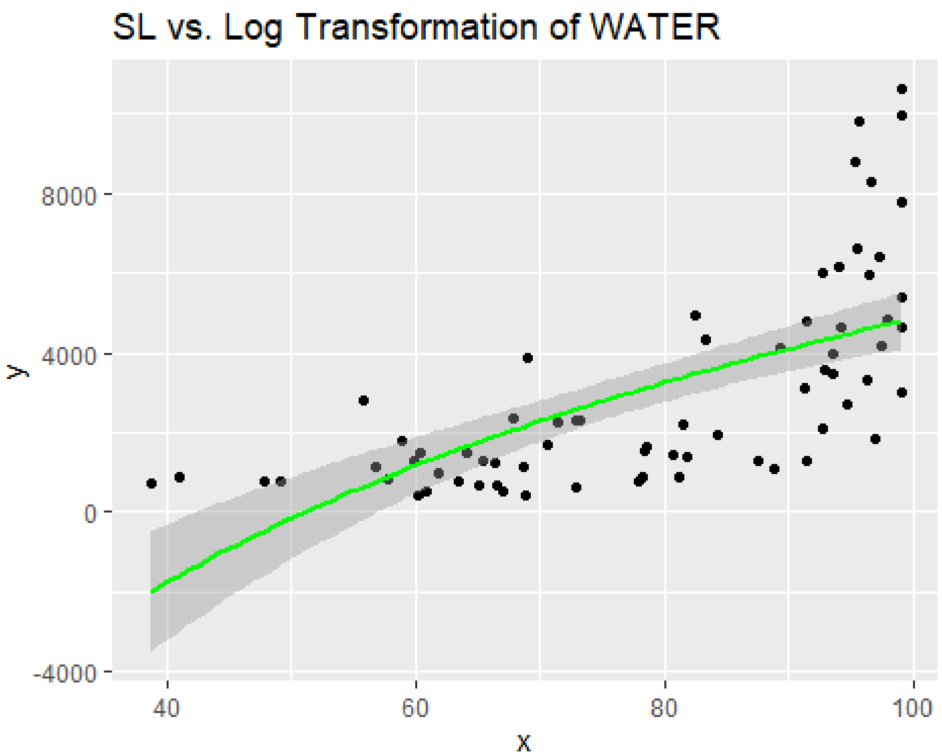



```
# From graphs, I can conclude that the cubic transformation is the best
# -----------------------------------------------------------------------

# -----------------------------------------------------------------------
x = CASES

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.0587 0.8093

# Since p-value = 0.8093 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F    Pr(>F)
## 1     66
```

```
## 2     68 -2 8.3099 0.0006042 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Since p-value = 0.0006042 < 0.05, Cubic model is better

# Since CASES takes value of 0, I cannot have a log transformation.

# Hence, I can conclude that the cubic transformation is the best
# -------------------------------------------------------------------------

# -------------------------------------------------------------------------
x = WATER

linear = lm(y ~ x, data = data)
squared = lm(y ~ x + I(x^2), data = data)
cubed = lm(y ~ x + I(x^2) + I(x^3), data = data)


waldtest(squared, c("I(x^2)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2)
## Model 2: y ~ x
##   Res.Df Df      F Pr(>F)
## 1     67
## 2     68 -1 0.1376 0.7119

# Since p-value = 0.7119 > 0.05, Linear model is better

waldtest(cubed, c("I(x^2)", "I(x^3)"), vcov = vcovHC(cubed, "HC1"))

## Wald test
##
## Model 1: y ~ x + I(x^2) + I(x^3)
## Model 2: y ~ x
##   Res.Df Df      F    Pr(>F)
## 1     66
## 2     68 -2 15.354 3.348e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Since p-value = 3.348e-06 < 0.05, Cubic model is better

p = ggplot(data, aes(x = x, y = y)) + geom_point()
p + stat_smooth(method = "lm", formula = y ~ x + I(x^2) + I(x^3), color =
"red") + ggtitle("SL vs. Cubic Transformation of WATER")
```




```
p + stat_smooth(method = "lm", formula = y ~ log(x), color = "green") + gg
title("SL vs. Log Transformation of WATER")
```




```
# From graphs, I can conclude that the cubic transformation is the best
# ----------------------------------------------------------------------
```

D. Equations of Model Specifications

D.1. For Response CASES

$$CASES = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times SL) + (\beta_8 \times HAQI) + \epsilon \qquad (1.1)$$

$$log(CASES) = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times SL) + (\beta_8 \times HAQI) + \epsilon \qquad (1.2)$$

$$\begin{aligned} CASES = {} & \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) \\ & + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times SL) + (\beta_8 \times HAQI) \\ & + (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) + (\beta_{11} \times HCI \times HYGIENE) \\ & + (\beta_{12} \times HCI \times SANITATION) + (\beta_{13} \times HCI \times WATER) \\ & + (\beta_{14} \times HCI \times SL) + (\beta_{15} \times HCI \times HAQI) + (\beta_{16} \times MEAT \times PD) \\ & + (\beta_{17} \times MEAT \times HYGIENE) + (\beta_{18} \times MEAT \times SANITATION) \\ & + (\beta_{19} \times MEAT \times WATER) + (\beta_{20} \times MEAT \times SL) \\ & + (\beta_{21} \times MEAT \times HAQI) + (\beta_{22} \times PD \times HYGIENE) \\ & + (\beta_{23} \times PD \times SANITATION) + (\beta_{24} \times PD \times WATER) + (\beta_{25} \times PD \times SL) \\ & + (\beta_{26} \times PD \times HAQI) + (\beta_{27} \times HYGIENE \times SANITATION) \\ & + (\beta_{28} \times HYGIENE \times WATER) + (\beta_{29} \times HYGIENE \times SL) \\ & + (\beta_{30} \times HYGIENE \times HAQI) + (\beta_{31} \times SANITATION \times WATER) \\ & + (\beta_{32} \times SANITATION \times SL) + (\beta_{33} \times SANITATION \times HAQI) \\ & + (\beta_{34} \times WATER \times SL) + (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times SL \times HAQI) \\ & + \epsilon \end{aligned} \qquad (1.3)$$

$$
\begin{aligned}
log(CASES) = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) \\
&+ (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times SL) + (\beta_8 \times HAQI) \\
&+ (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) + (\beta_{11} \times HCI \times HYGIENE) \\
&+ (\beta_{12} \times HCI \times SANITATION) + (\beta_{13} \times HCI \times WATER) \\
&+ (\beta_{14} \times HCI \times SL) + (\beta_{15} \times HCI \times HAQI) + (\beta_{16} \times MEAT \times PD) \\
&+ (\beta_{17} \times MEAT \times HYGIENE) + (\beta_{18} \times MEAT \times SANITATION) \\
&+ (\beta_{19} \times MEAT \times WATER) + (\beta_{20} \times MEAT \times SL) \\
&+ (\beta_{21} \times MEAT \times HAQI) + (\beta_{22} \times PD \times HYGIENE) \\
&+ (\beta_{23} \times PD \times SANITATION) + (\beta_{24} \times PD \times WATER) + (\beta_{25} \times PD \times SL) \\
&+ (\beta_{26} \times PD \times HAQI) + (\beta_{27} \times HYGIENE \times SANITATION) \\
&+ (\beta_{28} \times HYGIENE \times WATER) + (\beta_{29} \times HYGIENE \times SL) \\
&+ (\beta_{30} \times HYGIENE \times HAQI) + (\beta_{31} \times SANITATION \times WATER) \\
&+ (\beta_{32} \times SANITATION \times SL) + (\beta_{33} \times SANITATION \times HAQI) \\
&+ (\beta_{34} \times WATER \times SL) + (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times SL \times HAQI) \\
&+ \epsilon \qquad (1.4)
\end{aligned}
$$

$$
\begin{aligned}
CASES = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) \\
&+ (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times SL) + (\beta_8 \times HAQI) \\
&+ (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) + (\beta_{11} \times HCI \times HYGIENE) \\
&+ (\beta_{12} \times HCI \times SANITATION) + (\beta_{13} \times HCI \times WATER) \\
&+ (\beta_{14} \times HCI \times SL) + (\beta_{15} \times HCI \times HAQI) + (\beta_{16} \times MEAT \times PD) \\
&+ (\beta_{17} \times MEAT \times HYGIENE) + (\beta_{18} \times MEAT \times SANITATION) \\
&+ (\beta_{19} \times MEAT \times WATER) + (\beta_{20} \times MEAT \times SL) \\
&+ (\beta_{21} \times MEAT \times HAQI) + (\beta_{22} \times PD \times HYGIENE) \\
&+ (\beta_{23} \times PD \times SANITATION) + (\beta_{24} \times PD \times WATER) + (\beta_{25} \times PD \times SL) \\
&+ (\beta_{26} \times PD \times HAQI) + (\beta_{27} \times HYGIENE \times SANITATION) \\
&+ (\beta_{28} \times HYGIENE \times WATER) + (\beta_{29} \times HYGIENE \times SL) \\
&+ (\beta_{30} \times HYGIENE \times HAQI) + (\beta_{31} \times SANITATION \times WATER) \\
&+ (\beta_{32} \times SANITATION \times SL) + (\beta_{33} \times SANITATION \times HAQI) \\
&+ (\beta_{34} \times WATER \times SL) + (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times SL \times HAQI) \\
&+ (\beta_{37} \times SANITATION^2) + (\beta_{38} \times SANITATION^3) + (\beta_{39} \times \log(SL)) \\
&+ (\beta_{40} \times WATER^2) + (\beta_{41} \times WATER^3) + \epsilon \qquad (1.5)
\end{aligned}
$$

$$
\begin{aligned}
log(CASES) = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) \\
&+ (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times SL) + (\beta_8 \times HAQI) \\
&+ (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) + (\beta_{11} \times HCI \times HYGIENE) \\
&+ (\beta_{12} \times HCI \times SANITATION) + (\beta_{13} \times HCI \times WATER) \\
&+ (\beta_{14} \times HCI \times SL) + (\beta_{15} \times HCI \times HAQI) + (\beta_{16} \times MEAT \times PD) \\
&+ (\beta_{17} \times MEAT \times HYGIENE) + (\beta_{18} \times MEAT \times SANITATION) \\
&+ (\beta_{19} \times MEAT \times WATER) + (\beta_{20} \times MEAT \times SL) \\
&+ (\beta_{21} \times MEAT \times HAQI) + (\beta_{22} \times PD \times HYGIENE) \\
&+ (\beta_{23} \times PD \times SANITATION) + (\beta_{24} \times PD \times WATER) + (\beta_{25} \times PD \times SL) \\
&+ (\beta_{26} \times PD \times HAQI) + (\beta_{27} \times HYGIENE \times SANITATION) \\
&+ (\beta_{28} \times HYGIENE \times WATER) + (\beta_{29} \times HYGIENE \times SL) \\
&+ (\beta_{30} \times HYGIENE \times HAQI) + (\beta_{31} \times SANITATION \times WATER) \\
&+ (\beta_{32} \times SANITATION \times SL) + (\beta_{33} \times SANITATION \times HAQI) \\
&+ (\beta_{34} \times WATER \times SL) + (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times SL \times HAQI) \\
&+ (\beta_{37} \times SANITATION^2) + (\beta_{38} \times SANITATION^3) + (\beta_{39} \times \log(SL)) \\
&+ (\beta_{40} \times WATER^2) + (\beta_{41} \times WATER^3) + \epsilon \qquad (1.6)
\end{aligned}
$$

$$
\begin{aligned}
CASES = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3) \\
&+ (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3) \\
&+ (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times \log(PD)) \\
&+ (\beta_{11} \times SL) + (\beta_{12} \times HAQI) + (\beta_{13} \times MEAT \times PD) \\
&+ (\beta_{14} \times MEAT \times HAQI) + (\beta_{15} \times PD \times HYGIENE) \\
&+ (\beta_{16} \times PD \times SANITATION) + (\beta_{17} \times PD \times HAQI) \\
&+ (\beta_{18} \times WATER \times SL) + \epsilon \qquad (1.7)
\end{aligned}
$$

$$log(CASES) = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3)$$
$$+ (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3)$$
$$+ (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times \log(PD))$$
$$+ (\beta_{11} \times SL) + (\beta_{12} \times HAQI) + (\beta_{13} \times MEAT \times PD)$$
$$+ (\beta_{14} \times MEAT \times HAQI) + (\beta_{15} \times PD \times HYGIENE)$$
$$+ (\beta_{16} \times PD \times SANITATION) + (\beta_{17} \times PD \times HAQI)$$
$$+ (\beta_{18} \times WATER \times SL) + \epsilon \quad (1.8)$$

$$CASES = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3)$$
$$+ (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3)$$
$$+ (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times \log(PD))$$
$$+ (\beta_{11} \times SL) + (\beta_{12} \times HAQI) + \epsilon \quad (1.9)$$

$$log(CASES) = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3)$$
$$+ (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3)$$
$$+ (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times \log(PD))$$
$$+ (\beta_{11} \times SL) + (\beta_{12} \times HAQI) + \epsilon \quad (1.10)$$

$$
\begin{aligned}
CASES^{0.2121212} \\
&= \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3) \\
&+ (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3) \\
&+ (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times \log{(PD)}) \\
&+ (\beta_{11} \times SL) + (\beta_{12} \times HAQI) + (\beta_{13} \times MEAT \times PD) \\
&+ (\beta_{14} \times MEAT \times HAQI) + (\beta_{15} \times PD \times HYGIENE) \\
&+ (\beta_{16} \times PD \times SANITATION) + (\beta_{17} \times PD \times HAQI) \\
&+ (\beta_{18} \times WATER \times SL) + \epsilon \qquad (1.11)
\end{aligned}
$$

D.2. For Response SL

$$
\begin{aligned}
SL = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) \\
&+ (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times CASES) + (\beta_8 \times HAQI) + \epsilon \qquad (2.1)
\end{aligned}
$$

$$
\begin{aligned}
log(SL) = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) \\
&+ (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times CASES) \\
&+ (\beta_8 \times HAQI) + \epsilon \qquad (2.2)
\end{aligned}
$$

$$
\begin{aligned}
SL = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) \\
&+ (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times CASES) + (\beta_8 \times HAQI) \\
&+ (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) + (\beta_{11} \times HCI \times HYGIENE) \\
&+ (\beta_{12} \times HCI \times SANITATION) + (\beta_{13} \times HCI \times WATER) \\
&+ (\beta_{14} \times HCI \times CASES) + (\beta_{15} \times HCI \times HAQI) + (\beta_{16} \times MEAT \times PD) \\
&+ (\beta_{17} \times MEAT \times HYGIENE) + (\beta_{18} \times MEAT \times SANITATION) \\
&+ (\beta_{19} \times MEAT \times WATER) + (\beta_{20} \times MEAT \times CASES) \\
&+ (\beta_{21} \times MEAT \times HAQI) + (\beta_{22} \times PD \times HYGIENE) \\
&+ (\beta_{23} \times PD \times SANITATION) + (\beta_{24} \times PD \times WATER) \\
&+ (\beta_{25} \times PD \times CASES) + (\beta_{26} \times PD \times HAQI) \\
&+ (\beta_{27} \times HYGIENE \times SANITATION) + (\beta_{28} \times HYGIENE \times WATER) \\
&+ (\beta_{29} \times HYGIENE \times CASES) + (\beta_{30} \times HYGIENE \times HAQI) \\
&+ (\beta_{31} \times SANITATION \times WATER) + (\beta_{32} \times SANITATION \times CASES) \\
&+ (\beta_{33} \times SANITATION \times HAQI) + (\beta_{34} \times WATER \times CASES) \\
&+ (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times CASES \times HAQI) + \epsilon \qquad (2.3)
\end{aligned}
$$

$$
\begin{aligned}
log(SL) = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) \\
&+ (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times CASES) \\
&+ (\beta_8 \times HAQI) + (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) \\
&+ (\beta_{11} \times HCI \times HYGIENE) + (\beta_{12} \times HCI \times SANITATION) \\
&+ (\beta_{13} \times HCI \times WATER) + (\beta_{14} \times HCI \times CASES) + (\beta_{15} \times HCI \times HAQI) \\
&+ (\beta_{16} \times MEAT \times PD) + (\beta_{17} \times MEAT \times HYGIENE) \\
&+ (\beta_{18} \times MEAT \times SANITATION) + (\beta_{19} \times MEAT \times WATER) \\
&+ (\beta_{20} \times MEAT \times CASES) + (\beta_{21} \times MEAT \times HAQI) \\
&+ (\beta_{22} \times PD \times HYGIENE) + (\beta_{23} \times PD \times SANITATION) \\
&+ (\beta_{24} \times PD \times WATER) + (\beta_{25} \times PD \times CASES) + (\beta_{26} \times PD \times HAQI) \\
&+ (\beta_{27} \times HYGIENE \times SANITATION) + (\beta_{28} \times HYGIENE \times WATER) \\
&+ (\beta_{29} \times HYGIENE \times CASES) + (\beta_{30} \times HYGIENE \times HAQI) \\
&+ (\beta_{31} \times SANITATION \times WATER) + (\beta_{32} \times SANITATION \times CASES) \\
&+ (\beta_{33} \times SANITATION \times HAQI) + (\beta_{34} \times WATER \times CASES) \\
&+ (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times CASES \times HAQI) + \epsilon \qquad (2.4)
\end{aligned}
$$

$$
\begin{aligned}
SL = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) + (\beta_4 \times HCI) \\
&+ (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times CASES) + (\beta_8 \times HAQI) \\
&+ (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) + (\beta_{11} \times HCI \times HYGIENE) \\
&+ (\beta_{12} \times HCI \times SANITATION) + (\beta_{13} \times HCI \times WATER) \\
&+ (\beta_{14} \times HCI \times CASES) + (\beta_{15} \times HCI \times HAQI) + (\beta_{16} \times MEAT \times PD) \\
&+ (\beta_{17} \times MEAT \times HYGIENE) + (\beta_{18} \times MEAT \times SANITATION) \\
&+ (\beta_{19} \times MEAT \times WATER) + (\beta_{20} \times MEAT \times CASES) \\
&+ (\beta_{21} \times MEAT \times HAQI) + (\beta_{22} \times PD \times HYGIENE) \\
&+ (\beta_{23} \times PD \times SANITATION) + (\beta_{24} \times PD \times WATER) \\
&+ (\beta_{25} \times PD \times CASES) + (\beta_{26} \times PD \times HAQI) \\
&+ (\beta_{27} \times HYGIENE \times SANITATION) + (\beta_{28} \times HYGIENE \times WATER) \\
&+ (\beta_{29} \times HYGIENE \times CASES) + (\beta_{30} \times HYGIENE \times HAQI) \\
&+ (\beta_{31} \times SANITATION \times WATER) + (\beta_{32} \times SANITATION \times CASES) \\
&+ (\beta_{33} \times SANITATION \times HAQI) + (\beta_{34} \times WATER \times CASES) \\
&+ (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times CASES \times HAQI) + (\beta_{37} \times WATER^2) \\
&+ (\beta_{38} \times SANITATION^2) + (\beta_{39} \times MEAT^2) + (\beta_{40} \times PD^2) \\
&+ (\beta_{41} \times HAQI^2) + (\beta_{42} \times WATER^3) + (\beta_{43} \times SANITATION^3) \\
&+ (\beta_{44} \times MEAT^3) + (\beta_{45} \times PD^3) + (\beta_{46} \times HAQI^3) + \epsilon \qquad (2.5)
\end{aligned}
$$

$$
\begin{aligned}
log(SL) = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times SANITATION) + (\beta_3 \times HYGIENE) \\
&+ (\beta_4 \times HCI) + (\beta_5 \times MEAT) + (\beta_6 \times PD) + (\beta_7 \times CASES) \\
&+ (\beta_8 \times HAQI) + (\beta_9 \times HCI \times MEAT) + (\beta_{10} \times HCI \times PD) \\
&+ (\beta_{11} \times HCI \times HYGIENE) + (\beta_{12} \times HCI \times SANITATION) \\
&+ (\beta_{13} \times HCI \times WATER) + (\beta_{14} \times HCI \times CASES) + (\beta_{15} \times HCI \times HAQI) \\
&+ (\beta_{16} \times MEAT \times PD) + (\beta_{17} \times MEAT \times HYGIENE) \\
&+ (\beta_{18} \times MEAT \times SANITATION) + (\beta_{19} \times MEAT \times WATER) \\
&+ (\beta_{20} \times MEAT \times CASES) + (\beta_{21} \times MEAT \times HAQI) \\
&+ (\beta_{22} \times PD \times HYGIENE) + (\beta_{23} \times PD \times SANITATION) \\
&+ (\beta_{24} \times PD \times WATER) + (\beta_{25} \times PD \times CASES) + (\beta_{26} \times PD \times HAQI) \\
&+ (\beta_{27} \times HYGIENE \times SANITATION) + (\beta_{28} \times HYGIENE \times WATER) \\
&+ (\beta_{29} \times HYGIENE \times CASES) + (\beta_{30} \times HYGIENE \times HAQI) \\
&+ (\beta_{31} \times SANITATION \times WATER) + (\beta_{32} \times SANITATION \times CASES) \\
&+ (\beta_{33} \times SANITATION \times HAQI) + (\beta_{34} \times WATER \times CASES) \\
&+ (\beta_{35} \times WATER \times HAQI) + (\beta_{36} \times CASES \times HAQI) + (\beta_{37} \times WATER^2) \\
&+ (\beta_{38} \times SANITATION^2) + (\beta_{39} \times MEAT^2) + (\beta_{40} \times PD^2) \\
&+ (\beta_{41} \times HAQI^2) + (\beta_{42} \times WATER^3) + (\beta_{43} \times SANITATION^3) \\
&+ (\beta_{44} \times MEAT^3) + (\beta_{45} \times PD^3) + (\beta_{46} \times HAQI^3) + \epsilon \qquad (2.6)
\end{aligned}
$$

$$SL = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3) + (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3) + (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times MEAT^2) + (\beta_{11} \times MEAT^3) + (\beta_{12} \times PD) + (\beta_{13} \times CASES) + (\beta_{14} \times HAQI) + (\beta_{15} \times PD^2) + (\beta_{16} \times PD^3) + (\beta_{17} \times HYGIENE \times CASES) + (\beta_{18} \times MEAT \times SANITATION) + (\beta_{19} \times HYGIENE \times SANITATION) + \epsilon \tag{2.7}$$

$$log(SL) = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3) + (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3) + (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times MEAT^2) + (\beta_{11} \times MEAT^3) + (\beta_{12} \times PD) + (\beta_{13} \times CASES) + (\beta_{14} \times HAQI) + (\beta_{15} \times PD^2) + (\beta_{16} \times PD^3) + (\beta_{17} \times HYGIENE \times CASES) + (\beta_{18} \times MEAT \times SANITATION) + (\beta_{19} \times HYGIENE \times SANITATION) + \epsilon \tag{2.8}$$

$$SL = \beta_0 + (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3) + (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3) + (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times MEAT^2) + (\beta_{11} \times MEAT^3) + (\beta_{12} \times PD) + (\beta_{13} \times CASES) + (\beta_{14} \times HAQI) + (\beta_{15} \times PD^2) + (\beta_{16} \times PD^3) + \epsilon \tag{2.9}$$

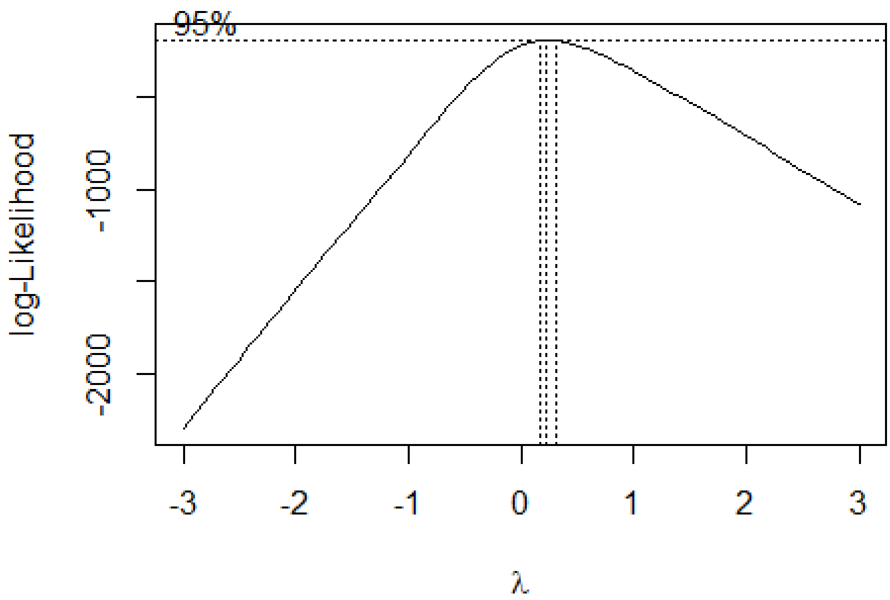
$$
\begin{aligned}
log(SL) = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3) \\
&+ (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3) \\
&+ (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times MEAT^2) \\
&+ (\beta_{11} \times MEAT^3) + (\beta_{12} \times PD) + (\beta_{13} \times CASES) + (\beta_{14} \times HAQI) \\
&+ (\beta_{15} \times PD^2) + (\beta_{16} \times PD^3) + \epsilon
\end{aligned} \tag{2.10}
$$

$$
\begin{aligned}
SL^{0.2727273} = \beta_0 &+ (\beta_1 \times WATER) + (\beta_2 \times WATER^2) + (\beta_3 \times WATER^3) \\
&+ (\beta_4 \times SANITATION) + (\beta_5 \times SANITATION^2) + (\beta_6 \times SANITATION^3) \\
&+ (\beta_7 \times HYGIENE) + (\beta_8 \times HCI) + (\beta_9 \times MEAT) + (\beta_{10} \times MEAT^2) \\
&+ (\beta_{11} \times MEAT^3) + (\beta_{12} \times PD) + (\beta_{13} \times CASES) + (\beta_{14} \times HAQI) \\
&+ (\beta_{15} \times PD^2) + (\beta_{16} \times PD^3) + (\beta_{17} \times HYGIENE \times CASES) \\
&+ (\beta_{18} \times MEAT \times SANITATION) + (\beta_{19} \times HYGIENE \times SANITATION) \\
&+ \epsilon
\end{aligned} \tag{2.11}
$$

E. Computation of different models for equation 1

```r
# Equation 1 Final Model

Model1 = lm(CASES ~ ., data = data)
summary(Model1)$adj.r.squared

## [1] 0.07054071

Model1 = lm(I(log(CASES)) ~ ., data = data)
summary(Model1)$adj.r.squared

## [1] 0.06985396

Model1 = lm(CASES ~ (.)^2, data = data)
summary(Model1)$adj.r.squared

## [1] 0.6110803

Model1 = lm(I(log(CASES)) ~ (.)^2, data = data)
summary(Model1)$adj.r.squared

## [1] 0.1097648
```

```
Model1 = lm(CASES ~ (.)^2 + I(SANITATION^2)+ I(SANITATION^3) + I(log(SL))
+ I(WATER^2) + I(WATER^3), data = data)
summary(Model1)$adj.r.squared

## [1] 0.5579806

Model1 = lm(I(log(CASES)) ~ (.)^2 + I(SANITATION^2)+ I(SANITATION^3) + I(l
og(SL)) + I(WATER^2) + I(WATER^3), data = data)
summary(Model1)$adj.r.squared

## [1] 0.04173011

# This One
Model1 = lm(CASES ~ (HCI + HAQI + HYGIENE + MEAT + PD + SANITATION + I(SAN
ITATION^2) + I(SANITATION^3) + I(log(SL)) + WATER + I(WATER^2) + I(WATER^3
) + MEAT*PD + MEAT*HAQI + PD*HYGIENE + PD*SANITATION + PD*HAQI + WATER*SL)
, data = data)
summary(Model1)$adj.r.squared

## [1] 0.6327776

Model1 = lm(I(log(CASES)) ~ (HCI + HAQI + HYGIENE + MEAT + PD + SANITATION
+ I(SANITATION^2) + I(SANITATION^3) + I(log(SL)) + WATER + I(WATER^2) + I(
WATER^3) + MEAT*PD + MEAT*HAQI + PD*HYGIENE + PD*SANITATION + PD*HAQI + WA
TER*SL), data = data)
summary(Model1)$adj.r.squared

## [1] 0.1636437

Model1 = lm(CASES ~ HCI + HAQI + HYGIENE + MEAT + PD + SANITATION + I(SANI
TATION^2) + I(SANITATION^3) + I(log(SL)) + WATER + I(WATER^2) + I(WATER^3)
, data = data)
summary(Model1)$adj.r.squared

## [1] 0.09639474

Model1 = lm(I(log(CASES)) ~ HCI + HAQI + HYGIENE + MEAT + PD + SANITATION
+ I(SANITATION^2) + I(SANITATION^3) + I(log(SL)) + WATER + I(WATER^2) + I(
WATER^3), data = data)
summary(Model1)$adj.r.squared

## [1] 0.1591429

# -------------------------

Model = lm(CASES ~ (HCI + HAQI + HYGIENE + MEAT + PD + SANITATION + I(SANI
TATION^2) + I(SANITATION^3) + I(log(SL)) + WATER + I(WATER^2) + I(WATER^3)
+ MEAT*PD + MEAT*HAQI + PD*HYGIENE + PD*SANITATION + PD*HAQI + WATER*SL),
data = data)

bc = boxcox(Model, lambda = seq(-3,3))
```




```
lam = bc$x[which(bc$y==max(bc$y))]
lam

## [1] 0.2121212

Model_bc = lm(I(CASES^lam) ~ (HCI + HAQI + HYGIENE + MEAT + PD + SANITATIO
N + I(SANITATION^2) + I(SANITATION^3) + I(log(SL)) + WATER + I(WATER^2) +
I(WATER^3) + MEAT*PD + MEAT*HAQI + PD*HYGIENE + PD*SANITATION + PD*HAQI +
WATER*SL), data = data)
summary(Model_bc)$adj.r.squared

## [1] 0.4899571

# The boxcox transformation worsens our model's fit. Hence, I must not use
it.

FinalModel1 = lm(CASES ~ (HCI + HAQI + HYGIENE + MEAT + PD + SANITATION +
I(SANITATION^2) + I(SANITATION^3) + I(log(SL)) + WATER + I(WATER^2) + I(WA
TER^3) + MEAT*PD + MEAT*HAQI + PD*HYGIENE + PD*SANITATION + PD*HAQI + WATE
R*SL), data = data)
summary(FinalModel1)

##
## Call:
## lm(formula = CASES ~ (HCI + HAQI + HYGIENE + MEAT + PD + SANITATION +
##     I(SANITATION^2) + I(SANITATION^3) + I(log(SL)) + WATER +
##     I(WATER^2) + I(WATER^3) + MEAT * PD + MEAT * HAQI + PD *
##     HYGIENE + PD * SANITATION + PD * HAQI + WATER * SL), data = data)
##
## Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -3791.6  -624.2   -49.3   526.5  5498.6
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -8.466e+03  2.013e+04  -0.421  0.67587
## HCI               8.887e+00  4.166e+01   0.213  0.83193
## HAQI              1.062e+02  8.275e+01   1.283  0.20540
## HYGIENE          -4.588e+01  1.943e+01  -2.362  0.02211 *
## MEAT             -9.268e+01  1.157e+02  -0.801  0.42691
## PD                7.140e+01  9.337e+00   7.647 5.86e-10 ***
## SANITATION       -2.936e+02  1.616e+02  -1.817  0.07517 .
## I(SANITATION^2)  5.843e+00  3.299e+00   1.771  0.08258 .
## I(SANITATION^3) -3.528e-02  1.983e-02  -1.779  0.08128 .
## I(log(SL))       -1.476e+02  1.414e+03  -0.104  0.91729
## WATER             3.951e+02  9.066e+02   0.436  0.66488
## I(WATER^2)       -4.283e+00  1.314e+01  -0.326  0.74589
## I(WATER^3)        1.204e-02  6.304e-02   0.191  0.84936
## SL               -1.459e+00  2.599e+00  -0.561  0.57697
## MEAT:PD          -6.743e-01  1.187e-01  -5.678 6.89e-07 ***
## HAQI:MEAT         1.524e+00  2.131e+00   0.715  0.47770
## HYGIENE:PD        5.799e-01  9.573e-02   6.058 1.78e-07 ***
## PD:SANITATION     4.159e-01  1.251e-01   3.324  0.00167 **
## HAQI:PD          -2.088e+00  2.590e-01  -8.061 1.34e-10 ***
## WATER:SL          2.174e-02  2.478e-02   0.877  0.38445
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1603 on 50 degrees of freedom
## Multiple R-squared:  0.7339, Adjusted R-squared:  0.6328
## F-statistic: 7.258 on 19 and 50 DF,  p-value: 9.406e-09
```

```
hetero_FinalModel1 = coeftest(FinalModel1, vcov = vcovHC(FinalModel1, "HC1
"))
hetero_FinalModel1
```

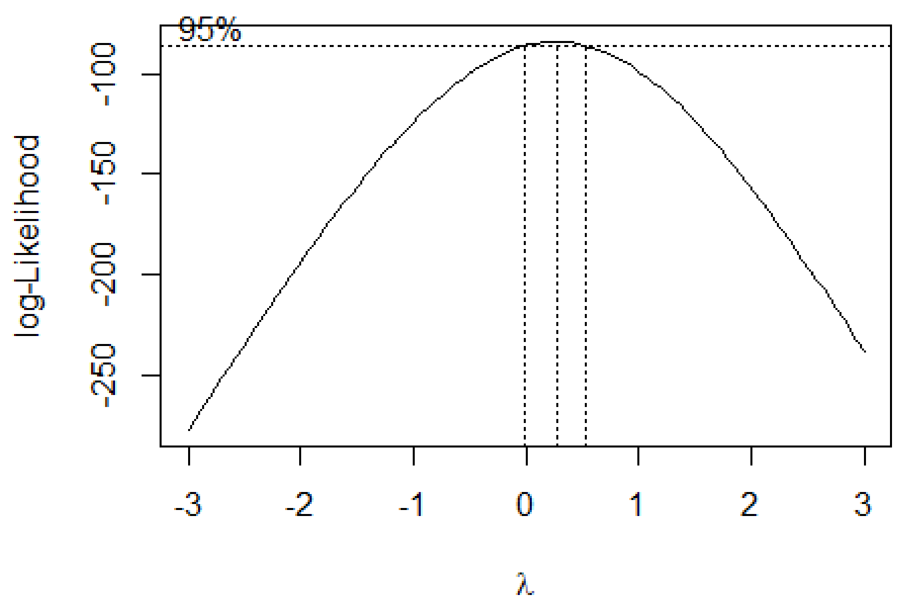
```
##
## t test of coefficients:
##
##                    Estimate  Std. Error t value Pr(>|t|)
## (Intercept)      -8.4656e+03  1.5958e+04 -0.5305 0.598122
## HCI               8.8868e+00  3.7795e+01  0.2351 0.815067
## HAQI              1.0618e+02  8.1286e+01  1.3062 0.197456
## HYGIENE          -4.5883e+01  2.0306e+01 -2.2597 0.028232 *
## MEAT             -9.2681e+01  1.0781e+02 -0.8597 0.394067
## PD                7.1403e+01  2.1773e+01  3.2794 0.001898 **
## SANITATION       -2.9359e+02  1.3490e+02 -2.1763 0.034284 *
## I(SANITATION^2)  5.8435e+00  2.8010e+00  2.0862 0.042084 *
## I(SANITATION^3) -3.5285e-02  1.7327e-02 -2.0364 0.047026 *
## I(log(SL))       -1.4759e+02  1.7759e+03 -0.0831 0.934095
## WATER             3.9508e+02  6.3717e+02  0.6201 0.538037
## I(WATER^2)       -4.2826e+00  9.3474e+00 -0.4582 0.648824
## I(WATER^3)        1.2035e-02  4.5736e-02  0.2631 0.793522
## SL               -1.4595e+00  2.7708e+00 -0.5267 0.600709
## MEAT:PD          -6.7434e-01  2.4633e-01 -2.7375 0.008551 **
```

```
## HAQI:MEAT        1.5244e+00  1.9872e+00  0.7671 0.446627
## HYGIENE:PD       5.7993e-01  1.7364e-01  3.3398 0.001591 **
## PD:SANITATION    4.1591e-01  1.5832e-01  2.6270 0.011406 *
## HAQI:PD         -2.0875e+00  6.4407e-01 -3.2411 0.002121 **
## WATER:SL         2.1744e-02  2.6096e-02  0.8332 0.408671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F. Computation of different models for equation 2

```
# Equation 2 Models

Model2 = lm(SL ~ ., data = data)
summary(Model2)$adj.r.squared

## [1] 0.6267465

Model2 = lm(I(log(SL)) ~ ., data = data)
summary(Model2)$adj.r.squared

## [1] 0.6462311

Model2 = lm(SL ~ (.)^2, data = data)
summary(Model2)$adj.r.squared

## [1] 0.7418093

Model2 = lm(I(log(SL)) ~ (.)^2, data = data)
summary(Model2)$adj.r.squared

## [1] 0.7042118

Model2 = lm(SL ~ (.)^2 + I(HAQI^2) + I(HAQI^3) + I(MEAT^2)+ I(MEAT^3) + I(
PD^2) + I(PD^3) + I(SANITATION^2) + I(SANITATION^3) + I(WATER^2) + I(WATER
^3), data = data)
summary(Model2)$adj.r.squared

## [1] 0.7365284

Model2 = lm(I(log(SL)) ~ (.)^2 + I(HAQI^2) + I(HAQI^3) + I(MEAT^2)+ I(MEAT
^3) + I(PD^2) + I(PD^3) + I(SANITATION^2) + I(SANITATION^3) + I(WATER^2) +
I(WATER^3), data = data)
summary(Model2)$adj.r.squared

## [1] 0.7538503

# This One
Model2 = lm(SL ~ HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + MEAT + I(M
EAT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2)
+ I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) + CASES*HYGIENE + MEAT
*SANITATION + HYGIENE*SANITATION, data = data)
summary(Model2)$adj.r.squared

## [1] 0.7679294
```

```
Model2 = lm(I(log(SL)) ~ HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + ME
AT + I(MEAT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION + I(SANIT
ATION^2) + I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) + CASES*HYGIE
NE + MEAT*SANITATION + HYGIENE*SANITATION, data = data)
summary(Model2)$adj.r.squared

## [1] 0.6802257

Model2 = lm(SL ~ (HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + MEAT + I(
MEAT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2
) + I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3)), data = data)
summary(Model2)$adj.r.squared

## [1] 0.7222186

Model2 = lm(I(log(SL)) ~ (HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + M
EAT + I(MEAT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION + I(SANI
TATION^2) + I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3)), data = dat
a)
summary(Model2)$adj.r.squared

## [1] 0.6653579

# -------------------------

Model = lm(SL ~ HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + MEAT + I(ME
AT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2)
+ I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) + CASES*HYGIENE + MEAT
*SANITATION + HYGIENE*SANITATION, data = data)
bc = boxcox(Model, lambda = seq(-3,3))
```



```
lam = bc$x[which(bc$y==max(bc$y))]
lam

## [1] 0.2727273

Model_bc = lm(I(SL^lam) ~ HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + M
EAT + I(MEAT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION + I(SANI
TATION^2) + I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) + CASES*HYGI
ENE + MEAT*SANITATION + HYGIENE*SANITATION, data = data)
summary(Model_bc)$adj.r.squared

## [1] 0.7176586

# The boxcox transformation worsens our model's fit. Hence, I must not use
it.

FinalModel2_1 = lm(SL ~ HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + MEA
T + I(MEAT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION + I(SANITA
TION^2) + I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) + CASES*HYGIEN
E + MEAT*SANITATION + HYGIENE*SANITATION, data = data)
summary(FinalModel2_1)

##
## Call:
## lm(formula = SL ~ HCI + HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE +
##     MEAT + I(MEAT^2) + I(MEAT^3) + PD + I(PD^2) + I(PD^3) + SANITATION
+
##     I(SANITATION^2) + I(SANITATION^3) + WATER + I(WATER^2) +
##     I(WATER^3) + CASES * HYGIENE + MEAT * SANITATION + HYGIENE *
##     SANITATION, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2082.05  -726.77   -28.07   589.19  2825.98
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.951e+04  2.615e+04  -1.128   0.2648
## HCI               1.335e+01  3.087e+01   0.433   0.6672
## HAQI              2.007e+03  1.269e+03   1.582   0.1202
## I(HAQI^2)        -4.029e+01  2.458e+01  -1.639   0.1077
## I(HAQI^3)         2.602e-01  1.541e-01   1.689   0.0977 .
## HYGIENE           6.128e+01  3.181e+01   1.926   0.0600 .
## MEAT              3.715e+01  9.728e+01   0.382   0.7042
## I(MEAT^2)        -2.811e+00  3.400e+00  -0.827   0.4125
## I(MEAT^3)         1.978e-02  2.519e-02   0.785   0.4363
## PD                2.078e+00  4.710e+00   0.441   0.6611
## I(PD^2)          -1.621e-02  1.279e-02  -1.267   0.2111
## I(PD^3)           1.151e-05  7.220e-06   1.595   0.1174
## SANITATION        7.984e+01  1.275e+02   0.626   0.5343
## I(SANITATION^2)  -2.039e+00  2.583e+00  -0.789   0.4337
## I(SANITATION^3)   1.785e-02  1.552e-02   1.150   0.2559
## WATER            -8.052e+01  6.853e+02  -0.117   0.9070
## I(WATER^2)       -5.841e-01  9.885e+00  -0.059   0.9531
## I(WATER^3)        1.236e-02  4.644e-02   0.266   0.7913
```

```
## CASES                -8.501e-01  4.268e-01  -1.992   0.0521 .
## HYGIENE:CASES         1.561e-02  6.623e-03   2.357   0.0225 *
## MEAT:SANITATION       1.865e+00  7.765e-01   2.401   0.0203 *
## HYGIENE:SANITATION -1.289e+00  5.685e-01  -2.268   0.0279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1239 on 48 degrees of freedom
## Multiple R-squared:  0.8386, Adjusted R-squared:  0.7679
## F-statistic: 11.87 on 21 and 48 DF,  p-value: 1.427e-12
```

```
hetero_FinalModel2_1 = coeftest(FinalModel2_1, vcov = vcovHC(FinalModel2_1
, "HC1"))
hetero_FinalModel2_1
```

```
##
## t test of coefficients:
##
##                       Estimate  Std. Error t value Pr(>|t|)
## (Intercept)        -2.9506e+04  2.4468e+04 -1.2059 0.233765
## HCI                 1.3354e+01  3.1069e+01  0.4298 0.669257
## HAQI                2.0071e+03  1.2047e+03  1.6661 0.102208
## I(HAQI^2)          -4.0290e+01  2.4108e+01 -1.6713 0.101176
## I(HAQI^3)           2.6017e-01  1.5368e-01  1.6929 0.096954 .
## HYGIENE             6.1279e+01  2.8481e+01  2.1516 0.036490 *
## MEAT                3.7152e+01  1.0047e+02  0.3698 0.713167
## I(MEAT^2)          -2.8111e+00  3.6225e+00 -0.7760 0.441554
## I(MEAT^3)           1.9779e-02  2.6053e-02  0.7592 0.451454
## PD                  2.0776e+00  5.5118e+00  0.3769 0.707884
## I(PD^2)            -1.6205e-02  1.3405e-02 -1.2090 0.232602
## I(PD^3)             1.1513e-05  7.2012e-06  1.5988 0.116428
## SANITATION          7.9840e+01  1.2643e+02  0.6315 0.530700
## I(SANITATION^2)    -2.0389e+00  2.8846e+00 -0.7068 0.483091
## I(SANITATION^3)     1.7851e-02  1.8157e-02  0.9831 0.330464
## WATER              -8.0521e+01  6.6734e+02 -0.1207 0.904465
## I(WATER^2)         -5.8409e-01  1.0064e+01 -0.0580 0.953961
## I(WATER^3)          1.2360e-02  4.9172e-02  0.2514 0.802607
## CASES              -8.5007e-01  2.9551e-01 -2.8767 0.005979 **
## HYGIENE:CASES       1.5612e-02  4.8643e-03  3.2095 0.002372 **
## MEAT:SANITATION     1.8647e+00  8.4541e-01  2.2056 0.032232 *
## HYGIENE:SANITATION -1.2891e+00  6.3263e-01 -2.0376 0.047119 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

G. Computation of Instrument Relevance for Model 1

```
summary(lm(HYGIENE ~ HCI + HAQI + MEAT +
             PD + I(log(SL)) + SANITATION + I(SANITATION^2) + MEAT*PD +
             MEAT*HAQI + PD*HYGIENE + WATER + I(WATER^2) + PD*SANITATION +
             PD*HAQI + WATER*SL, data = data))
```

```
## Warning in model.matrix.default(mt, mf, contrasts): the response appear
ed on the
## right-hand side and was dropped

## Warning in model.matrix.default(mt, mf, contrasts): problem with term 8
in
## model.matrix: no columns are assigned

##
## Call:
## lm(formula = HYGIENE ~ HCI + HAQI + MEAT + PD + I(log(SL)) +
##     SANITATION + I(SANITATION^2) + MEAT * PD + MEAT * HAQI +
##     PD * HYGIENE + WATER + I(WATER^2) + PD * SANITATION + PD *
##     HAQI + WATER * SL, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.6154  -4.9025   0.8666   4.5101  20.7020
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.248e+02  7.562e+01  -1.650  0.10479
## HCI              -2.329e-02  2.783e-01  -0.084  0.93364
## HAQI              1.340e+00  5.425e-01   2.470  0.01675 *
## MEAT              1.180e+00  7.641e-01   1.544  0.12857
## PD                1.812e-01  6.087e-02   2.977  0.00439 **
## I(log(SL))        6.059e+00  9.270e+00   0.654  0.51618
## SANITATION        2.296e-01  3.742e-01   0.614  0.54215
## I(SANITATION^2)  4.012e-03  3.294e-03   1.218  0.22867
## WATER             1.249e+00  1.238e+00   1.009  0.31768
## I(WATER^2)       -1.045e-02  9.882e-03  -1.058  0.29507
## SL               -1.141e-02  1.690e-02  -0.675  0.50271
## MEAT:PD          -1.194e-03  7.910e-04  -1.509  0.13715
## HAQI:MEAT        -1.930e-02  1.413e-02  -1.365  0.17788
## HYGIENE:PD        3.178e-03  5.139e-04   6.184 9.23e-08 ***
## PD:SANITATION    -1.615e-03  8.380e-04  -1.927  0.05931 .
## HAQI:PD          -4.060e-03  1.719e-03  -2.362  0.02188 *
## WATER:SL          1.133e-04  1.607e-04   0.705  0.48418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.35 on 53 degrees of freedom
## Multiple R-squared:  0.8924, Adjusted R-squared:  0.8599
## F-statistic: 27.47 on 16 and 53 DF,  p-value: < 2.2e-16
```

```
summary(lm(I(SANITATION^3) ~ HCI + HAQI + MEAT +
             PD + I(log(SL)) + SANITATION + I(SANITATION^2) + MEAT*PD +
             MEAT*HAQI + PD*HYGIENE + WATER + I(WATER^2) + PD*SANITATION +
             PD*HAQI + WATER*SL, data = data))
```

```
##
## Call:
## lm(formula = I(SANITATION^3) ~ HCI + HAQI + MEAT + PD + I(log(SL)) +
##     SANITATION + I(SANITATION^2) + MEAT * PD + MEAT * HAQI +
##     PD * HYGIENE + WATER + I(WATER^2) + PD * SANITATION + PD *
```

```
##     HAQI + WATER * SL, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25970.1  -5827.0   -687.1   7700.8  24912.3
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -1.170e+05  8.305e+04  -1.409 0.164668
## HCI                 1.061e+02  2.981e+02   0.356 0.723394
## HAQI                1.216e+02  6.135e+02   0.198 0.843646
## MEAT               -2.843e+02  8.365e+02  -0.340 0.735376
## PD                 -4.941e+01  7.043e+01  -0.701 0.486154
## I(log(SL))          3.381e+03  9.968e+03   0.339 0.735838
## SANITATION         -7.556e+03  4.022e+02 -18.788  < 2e-16 ***
## I(SANITATION^2)     1.647e+02  3.577e+00  46.032  < 2e-16 ***
## HYGIENE            -3.411e+01  1.471e+02  -0.232 0.817553
## WATER               5.558e+03  1.339e+03   4.152 0.000123 ***
## I(WATER^2)         -4.081e+01  1.070e+01  -3.815 0.000363 ***
## SL                 -1.279e+01  1.818e+01  -0.703 0.484883
## MEAT:PD            -1.526e+00  8.651e-01  -1.764 0.083620 .
## HAQI:MEAT           1.291e+00  1.540e+01   0.084 0.933514
## PD:HYGIENE          4.814e-01  7.222e-01   0.667 0.507942
## PD:SANITATION      -1.216e+00  9.284e-01  -1.310 0.195935
## HAQI:PD             2.104e+00  1.935e+00   1.087 0.281874
## WATER:SL            1.522e-01  1.730e-01   0.880 0.383000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12150 on 52 degrees of freedom
## Multiple R-squared:  0.999,  Adjusted R-squared:  0.9986
## F-statistic:  2953 on 17 and 52 DF,  p-value: < 2.2e-16
```

```
summary(lm(I(WATER^3) ~ HCI + HAQI + MEAT +
             PD + I(log(SL)) + SANITATION + I(SANITATION^2) + MEAT*PD +
             MEAT*HAQI + PD*HYGIENE + WATER + I(WATER^2) + PD*SANITATION +
             PD*HAQI + WATER*SL, data = data))
```

```
##
## Call:
## lm(formula = I(WATER^3) ~ HCI + HAQI + MEAT + PD + I(log(SL)) +
##     SANITATION + I(SANITATION^2) + MEAT * PD + MEAT * HAQI +
##     PD * HYGIENE + WATER + I(WATER^2) + PD * SANITATION + PD *
##     HAQI + WATER * SL, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8967.1 -2328.4   213.3  2349.5  7706.0
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2.515e+05  2.613e+04   9.626 3.81e-13 ***
## HCI                -2.025e+02  9.379e+01  -2.159  0.03546 *
## HAQI               -2.557e+02  1.930e+02  -1.325  0.19111
```

```
## MEAT              -5.788e+02  2.632e+02  -2.199  0.03232 *
## PD                -6.481e+00  2.216e+01  -0.293  0.77107
## I(log(SL))         8.592e+03  3.136e+03   2.740  0.00840 **
## SANITATION         1.076e+02  1.265e+02   0.850  0.39898
## I(SANITATION^2)  2.622e-02  1.125e+00   0.023  0.98150
## HYGIENE           -1.054e+01  4.628e+01  -0.228  0.82068
## WATER             -1.334e+04  4.211e+02 -31.679  < 2e-16 ***
## I(WATER^2)         2.020e+02  3.365e+00  60.034  < 2e-16 ***
## SL                -1.698e+01  5.720e+00  -2.968  0.00452 **
## MEAT:PD            8.618e-02  2.722e-01   0.317  0.75280
## HAQI:MEAT          1.033e+01  4.845e+00   2.133  0.03768 *
## PD:HYGIENE         1.631e-02  2.272e-01   0.072  0.94303
## PD:SANITATION     -3.467e-01  2.921e-01  -1.187  0.24059
## HAQI:PD            5.630e-01  6.089e-01   0.925  0.35940
## WATER:SL           1.652e-01  5.441e-02   3.036  0.00374 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3823 on 52 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9998
## F-statistic: 2.364e+04 on 17 and 52 DF,  p-value: < 2.2e-16
```

H. Computation of Instrument Relevance for Model 2

```
summary(lm(HYGIENE ~ MEAT + I(MEAT^2) + I(MEAT^3) +
             I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2) + HAQI +
             I(HAQI^2) + I(HAQI^3) + WATER + I(WATER^2) + + CASES*HYGIENE
+
             MEAT*SANITATION + HYGIENE*SANITATION + PD + HCI, data = data)
)

## Warning in model.matrix.default(mt, mf, contrasts): the response appear
ed on the
## right-hand side and was dropped

## Warning in model.matrix.default(mt, mf, contrasts): problem with term 1
4 in
## model.matrix: no columns are assigned

##
## Call:
## lm(formula = HYGIENE ~ MEAT + I(MEAT^2) + I(MEAT^3) + I(PD^2) +
##     I(PD^3) + SANITATION + I(SANITATION^2) + HAQI + I(HAQI^2) +
##     I(HAQI^3) + WATER + I(WATER^2) + +CASES * HYGIENE + MEAT *
##     SANITATION + HYGIENE * SANITATION + PD + HCI, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.6281  -3.1354   0.3963   2.5403  16.2210
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)         -7.466e+01  9.323e+01  -0.801  0.42699    
## MEAT                -8.620e-01  4.113e-01  -2.096  0.04107 *  
## I(MEAT^2)            3.929e-02  1.376e-02   2.856  0.00620 ** 
## I(MEAT^3)           -2.896e-04  1.023e-04  -2.831  0.00662 ** 
## I(PD^2)              1.201e-05  5.434e-05   0.221  0.82592    
## I(PD^3)             -9.793e-09  3.102e-08  -0.316  0.75350    
## SANITATION           7.880e-01  1.660e-01   4.746 1.72e-05 ***
## I(SANITATION^2)     -1.238e-02  1.804e-03  -6.861 9.02e-09 ***
## HAQI                 3.448e+00  5.567e+00   0.619  0.53842    
## I(HAQI^2)           -5.771e-02  1.080e-01  -0.534  0.59537    
## I(HAQI^3)            3.360e-04  6.771e-04   0.496  0.62192    
## WATER                4.768e-01  5.275e-01   0.904  0.37038    
## I(WATER^2)          -3.402e-03  3.708e-03  -0.917  0.36322    
## CASES                4.372e-04  1.837e-03   0.238  0.81285    
## PD                  -3.147e-03  1.954e-02  -0.161  0.87271    
## HCI                 -4.802e-02  1.319e-01  -0.364  0.71727    
## HYGIENE:CASES       -1.027e-05  2.838e-05  -0.362  0.71909    
## MEAT:SANITATION     -8.683e-03  2.953e-03  -2.940  0.00492 ** 
## HYGIENE:SANITATION   1.662e-02  9.200e-04  18.060  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 5.459 on 51 degrees of freedom
## Multiple R-squared:  0.976,  Adjusted R-squared:  0.9676 
## F-statistic: 115.4 on 18 and 51 DF,  p-value: < 2.2e-16

summary(lm(I(SANITATION^3) ~ MEAT + I(MEAT^2) + I(MEAT^3) + 
             I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2) + HAQI + 
             I(HAQI^2) + I(HAQI^3) + WATER + I(WATER^2) + + CASES*HYGIENE 
+ 
             MEAT*SANITATION + HYGIENE*SANITATION + PD + HCI, data = data)
)

## 
## Call:
## lm(formula = I(SANITATION^3) ~ MEAT + I(MEAT^2) + I(MEAT^3) + 
##     I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2) + HAQI + 
##     I(HAQI^2) + I(HAQI^3) + WATER + I(WATER^2) + +CASES * HYGIENE + 
##     MEAT * SANITATION + HYGIENE * SANITATION + PD + HCI, data = data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -19714.8  -7637.0   -556.1   7041.9  23669.4 
## 
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)    
## (Intercept)         -3.671e+04  2.043e+05  -0.180 0.858104    
## MEAT                 1.197e+02  9.332e+02   0.128 0.898463    
## I(MEAT^2)           -2.524e+01  3.226e+01  -0.782 0.437743    
## I(MEAT^3)            1.761e-01  2.396e-01   0.735 0.465668    
## I(PD^2)              2.132e-01  1.184e-01   1.801 0.077784 .  
## I(PD^3)             -9.597e-05  6.761e-05  -1.420 0.161949    
## SANITATION          -7.642e+03  4.341e+02 -17.604  < 2e-16 ***
## I(SANITATION^2)      1.610e+02  5.447e+00  29.555  < 2e-16 ***
```

```
## HAQI                  -1.269e+03  1.217e+04  -0.104 0.917324
## I(HAQI^2)              6.145e+00  2.358e+02   0.026 0.979313
## I(HAQI^3)              1.292e-01  1.478e+00   0.087 0.930684
## WATER                  4.732e+03  1.158e+03   4.087 0.000158 ***
## I(WATER^2)            -3.212e+01  8.140e+00  -3.946 0.000248 ***
## CASES                 -4.922e+00  4.003e+00  -1.230 0.224591
## HYGIENE               -1.024e+02  3.049e+02  -0.336 0.738426
## PD                    -1.071e+02  4.257e+01  -2.515 0.015164 *
## HCI                    2.252e+02  2.875e+02   0.783 0.437157
## CASES:HYGIENE          7.893e-02  6.188e-02   1.276 0.208026
## MEAT:SANITATION        8.465e+00  6.954e+00   1.217 0.229202
## SANITATION:HYGIENE     1.827e+00  5.448e+00   0.335 0.738775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11890 on 50 degrees of freedom
## Multiple R-squared:  0.999,  Adjusted R-squared:  0.9987
## F-statistic:  2762 on 19 and 50 DF,  p-value: < 2.2e-16

summary(lm(I(WATER^3) ~ MEAT + I(MEAT^2) + I(MEAT^3) +
             I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2) + HAQI +
             I(HAQI^2) + I(HAQI^3) + WATER + I(WATER^2) + + CASES*HYGIENE
+
             MEAT*SANITATION + HYGIENE*SANITATION + PD + HCI, data = data)
)

##
## Call:
## lm(formula = I(WATER^3) ~ MEAT + I(MEAT^2) + I(MEAT^3) + I(PD^2) +
##     I(PD^3) + SANITATION + I(SANITATION^2) + HAQI + I(HAQI^2) +
##     I(HAQI^3) + WATER + I(WATER^2) + +CASES * HYGIENE + MEAT *
##     SANITATION + HYGIENE * SANITATION + PD + HCI, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10574.0  -2029.6    -98.4   2755.4   7771.7
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.230e+05  6.829e+04   4.730 1.88e-05 ***
## MEAT                   5.590e+00  3.120e+02   0.018  0.98577
## I(MEAT^2)             -1.002e+01  1.078e+01  -0.929  0.35712
## I(MEAT^3)              6.489e-02  8.008e-02   0.810  0.42159
## I(PD^2)                4.769e-02  3.957e-02   1.205  0.23387
## I(PD^3)               -2.502e-05  2.260e-05  -1.107  0.27358
## SANITATION             4.105e+01  1.451e+02   0.283  0.77840
## I(SANITATION^2)       -1.367e+00  1.821e+00  -0.751  0.45635
## HAQI                  -8.593e+02  4.067e+03  -0.211  0.83352
## I(HAQI^2)              9.279e+00  7.882e+01   0.118  0.90676
## I(HAQI^3)              7.381e-03  4.941e-01   0.015  0.98814
## WATER                 -1.395e+04  3.870e+02 -36.046  < 2e-16 ***
## I(WATER^2)             2.085e+02  2.721e+00  76.614  < 2e-16 ***
## CASES                 -1.676e+00  1.338e+00  -1.252  0.21622
## HYGIENE               -1.757e+01  1.019e+02  -0.172  0.86380
```

```
## PD                  -1.372e+01  1.423e+01  -0.965  0.33941
## HCI                 -1.188e+02  9.612e+01  -1.236  0.22223
## CASES:HYGIENE        2.898e-02  2.069e-02   1.401  0.16745
## MEAT:SANITATION      6.255e+00  2.325e+00   2.691  0.00967 **
## SANITATION:HYGIENE -1.226e-03  1.821e+00  -0.001  0.99947
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3973 on 50 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9998
## F-statistic: 1.958e+04 on 19 and 50 DF,  p-value: < 2.2e-16
```

I. Results of IV Regression.

```
iv_1 = ivreg(CASES ~ HYGIENE + SANITATION + I(SANITATION^2) +
               I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) +
               PD + MEAT*PD + MEAT*HAQI + PD*HYGIENE +
               PD*SANITATION + PD*HAQI + WATER*SL | HCI + HAQI + MEAT +
               PD + I(log(SL)) + SANITATION + I(SANITATION^2) + MEAT*PD +
               MEAT*HAQI + PD*HYGIENE + WATER + I(WATER^2) + PD*SANITATION
+
               PD*HAQI + WATER*SL, data = data)
summary(iv_1, vcov = sandwich, df = Inf)

##
## Call:
## ivreg(formula = CASES ~ HYGIENE + SANITATION + I(SANITATION^2) +
##     I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) + PD +
##     MEAT * PD + MEAT * HAQI + PD * HYGIENE + PD * SANITATION +
##     PD * HAQI + WATER * SL | HCI + HAQI + MEAT + PD + I(log(SL)) +
##     SANITATION + I(SANITATION^2) + MEAT * PD + MEAT * HAQI +
##     PD * HYGIENE + WATER + I(WATER^2) + PD * SANITATION + PD *
##     HAQI + WATER * SL, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3914.5  -670.9  -170.7   627.7  5633.7
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        2.143e+03  3.545e+04   0.060  0.95180
## HYGIENE           -4.519e+01  1.801e+01  -2.509  0.01210 *
## SANITATION        -7.054e+01  1.704e+03  -0.041  0.96698
## I(SANITATION^2)    1.051e+00  3.715e+01   0.028  0.97743
## I(SANITATION^3)   -6.173e-03  2.260e-01  -0.027  0.97821
## WATER             -1.487e+02  1.751e+03  -0.085  0.93232
## I(WATER^2)         2.689e+00  2.338e+01   0.115  0.90841
## I(WATER^3)        -1.660e-02  1.160e-01  -0.143  0.88618
## PD                 7.266e+01  2.235e+01   3.250  0.00115 **
## MEAT              -1.010e+02  1.389e+02  -0.727  0.46720
## HAQI               9.532e+01  6.583e+01   1.448  0.14763
## SL                -1.573e+00  2.122e+00  -0.741  0.45844
```

```
## PD:MEAT          -6.274e-01  3.826e-01  -1.640  0.10100
## MEAT:HAQI         1.783e+00  1.875e+00   0.951  0.34170
## HYGIENE:PD        5.664e-01  1.820e-01   3.112  0.00186 **
## SANITATION:PD     4.414e-01  2.945e-01   1.499  0.13387
## PD:HAQI          -2.133e+00  7.271e-01  -2.933  0.00335 **
## WATER:SL          2.204e-02  2.807e-02   0.785  0.43231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1606 on Inf degrees of freedom
## Multiple R-Squared: 0.7224,  Adjusted R-squared: 0.6316
## Wald test: 36.82 on 17 DF,  p-value: 0.003559
```

```
iv_2 = ivreg(SL ~ HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE + SANITATION + I(
SANITATION^2) +
               I(SANITATION^3) + WATER + I(WATER^2) + I(WATER^3) + CASES*H
YGIENE +
               MEAT*SANITATION + HYGIENE*SANITATION | + MEAT + I(MEAT^2) +
I(MEAT^3) +
               I(PD^2) + I(PD^3) + SANITATION + I(SANITATION^2) + HAQI +
               I(HAQI^2) + I(HAQI^3) + WATER + I(WATER^2) + + CASES*HYGIEN
E +
               MEAT*SANITATION + HYGIENE*SANITATION , data = data)
summary(iv_2, vcov = sandwich, df = Inf)
```

```
##
## Call:
## ivreg(formula = SL ~ HAQI + I(HAQI^2) + I(HAQI^3) + HYGIENE +
##     SANITATION + I(SANITATION^2) + I(SANITATION^3) + WATER +
##     I(WATER^2) + I(WATER^3) + CASES * HYGIENE + MEAT * SANITATION +
##     HYGIENE * SANITATION | +MEAT + I(MEAT^2) + I(MEAT^3) + I(PD^2) +
##     I(PD^3) + SANITATION + I(SANITATION^2) + HAQI + I(HAQI^2) +
##     I(HAQI^3) + WATER + I(WATER^2) + +CASES * HYGIENE + MEAT *
##     SANITATION + HYGIENE * SANITATION, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3849.47 -1346.16   -66.23  1312.90  4199.25
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.464e+04  9.437e+04  -0.155   0.8767
## HAQI                  3.424e+03  1.411e+03   2.427   0.0152 *
## I(HAQI^2)            -6.646e+01  2.782e+01  -2.389   0.0169 *
## I(HAQI^3)             4.104e-01  1.762e-01   2.329   0.0199 *
## HYGIENE               7.042e+01  4.390e+01   1.604   0.1087
## SANITATION            1.262e+03  5.535e+02   2.280   0.0226 *
## I(SANITATION^2)      -2.702e+01  1.154e+01  -2.341   0.0192 *
## I(SANITATION^3)       1.718e-01  7.101e-02   2.419   0.0156 *
## WATER                -2.124e+03  4.473e+03  -0.475   0.6349
## I(WATER^2)            2.412e+01  6.434e+01   0.375   0.7078
## I(WATER^3)           -8.272e-02  2.988e-01  -0.277   0.7819
## CASES                -5.392e-01  7.016e-01  -0.768   0.4422
## MEAT                 -1.015e-01  7.824e+01  -0.001   0.9990
```

```
## HYGIENE:CASES       1.050e-02  1.107e-02   0.949   0.3428
## SANITATION:MEAT     1.223e+00  1.105e+00   1.107   0.2684
## HYGIENE:SANITATION -1.474e+00  8.101e-01  -1.819   0.0688 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2152 on Inf degrees of freedom
## Multiple R-Squared: 0.4519,  Adjusted R-squared: 0.2997
## Wald test: 283.7 on 15 DF,  p-value: < 2.2e-16
```