Yanni Guo

December 2021

Analysis of Heart Disease Mortality in Relation with Age, Gender, Oral and Nutritional Health Factors in the United States

**Abstract:**

This project aimed to analyze the relation and effects of oral health, nutritional health, age, and gender-related factors with cardiovascular disease mortality across the United States from 1999 to 2019. Data sets were collected from the Centers for Disease Control and Prevention (CDC), CDC Wonder, and the United States Department of Agriculture (USDA). I used R for data cleaning, wrangling, and creating models. Spline interpolation was used to fill in the missing data values. The fixed-effects models determined the significant predictors for heart disease mortality. A VIF test was completed to avoid multicollinearity. Normality plots and residual plots of the models were conducted to verify the model’s results.

Significant predictors for heart disease mortality were state levels, gender levels, number of teeth removed due to gum disease or cavities, number of dentist visits within the past year, food insecurity % per state, and number of farmer’s markets per state. Heart disease mortality decreased from 1999 to 2019. Males had higher numbers of heart disease deaths compared to females. Older age groups (65-84) had heart disease deaths compared to younger age groups (35-64). However, the significant predictors had weak correlations with heart disease mortality. The results implicated that other factors could be explored to study the association between the health risk factors and heart disease deaths.

**Introduction:**

Cardiovascular disease is a major leading cause of death in the United States. Due to the severity of cardiovascular disease, several studies have researched the disease’s risk factors to prevent its development (Roth, G. et al., 2020). The purpose of this project was to design a predictive model to analyze the trends and relations of heart disease mortality across the United States from 1999 to 2019 in relation to the following risk factors: oral and diet health, age, and gender. Fixed-effect linear regression models were created to determine the significant predictors (Midway, S., 2021). The variables I used to evaluate the specified risk factors were tooth removal (due to gum disease or cavities), dental visit in the past year, number of grocery stores, number of supercenters, number of convenience stores, number of fast-food restaurants, number of full-service restaurants, number of farmer’s market, and food insecurity %. I planned to answer these research questions: 1) how does heart disease mortality change geographically in the US between 1999 and 2019? 2) what variables can best predict heart disease death? 3) what is the pattern of heart disease mortality along with the pattern of the significant predictors? 4) what effects do each of the significant risk factors have on heart disease mortality in the US? I decided on these research questions because research that studied risk factor relations with heart disease mortality remained inconclusive ​​(Kim, K. et al., 2019; Bains, A. and Rashid, M., 2013). I was interested in studying how people’s oral health status has an association with heart disease because researchers have found a positive relationship between the two variables but are uncertain about the underlying relations. Hypothesis includes the deterioration of teeth health by excessive consumption of junk food or low consumption of healthy food (Batty, D et al., 2018). Furthermore, access to dental clinics and healthy and unhealthy food access depending on region and year may be important factors (Batty, D et al., 2018). Healthy and unhealthy food consumption are significant contributors to the health of people’s teeth (Moynihan, Paula, 2005). Moreover, researchers found that healthy diets, such as the Mediterranean diet rich in fruits and vegetables, can decrease the risk of heart disease death. An unhealthy diet, such as the Western diet, mainly composed of junk-food substances, increases the risk of heart disease death (Heidemann, C et al., 2008). But, researchers have not determined if people in age groups younger than 80 can benefit from a healthier diet and whether different percent consumption of food can affect heart disease death (Bhupathiraju, S., and Katherine, T., 2011). Lastly, determining the risk factors of heart disease is important because it will help health professions prevent heart disease in their patients and have more accurate predictions of heart disease. By understanding the risk factors, health professions will be able to bring more awareness to the public (Hemingway, H. et al., 2018). This research will also contribute to the translation of large volumes of data in health industries so the health professionals can provide high-quality care to their patients (Hemingway, H. et al., 201). Thus, I planned to fill in those knowledge gaps concerning heart disease death with my research questions and predictive model by fitting in the variables related to oral health, diet, junk food, and individual factors.

**Literature Review:**

**Oral Health domain:**

Oral health is often perceived as separate from other organs of the body, but researchers have found that oral health can be linked to the risk of cardiovascular heart disease. Studies have determined there exists an association between cardiovascular heart disease and oral health, more specifically tooth decay, and found that poor oral health leads to increased risk of heart disease (Moynihan, Paula, 2005; Kim, K. et al., 2019; Okoro, C. et al., 2005). The studies found variables that influenced the development of tooth decay were diet, nutrition, smoking, age, the interaction between smoking status and age group, gender, socioeconomic status, gender (Kim, K. et al., 2019). Males in middle-age groups were more common to have poor oral health than women in middle-age groups primarily due to the different consumption of alcohol and smoking (Batty, D et al., 2018). An unhealthy diet with a low intake of fruits and vegetables, sugar consumption, and fat consumption increased the occurrence of tooth decay which resulted in the risk increase of cardiovascular heart disease (Moynihan, Paula, 2005). Another interesting reason for the positive relation was that poor oral health caused tooth loss and inhibited the chewing of fruits, vegetables, and whole grains (Moynihan, Paula, 2005). However, researchers were uncertain if bacterial infection or inflammation in the root canal system may have interfered with the increased risk of heart disease in their patients. Several studies faced limitations including missing data and people facing barriers in receiving proper dental care due to cost (Batty, D et al., 2018). Thus, I used teeth removal (due to teeth cavities or gum disease) and dentist visits (within last year) to measure oral health. More studies on variable interactions and controlled variables can help dental clinics develop strategies to prevent tooth decay in their patients.

**Diet domain:**

A healthy diet and diet patterns are significant risk factors for cardiovascular disease. Studies determined two major diet patterns from food frequency questionnaires related to the development of heart disease (Heidemann, C et al., 2008; Aune, D et al., 2017). A healthy diet that includes a rich intake of fruits, vegetables, legumes, fish, poultry, and whole grains is an effective strategy to prevent cardiovascular heart disease. On the other hand, a non-healthy diet pattern (western pattern) with high consumption of red meat, processed meat, refined grains, french fries, and sweets/desserts lead to a higher risk of heart disease death (Heidemann, C et al., 2008). Several studies found that the Mediterranean diet or the Dietary Approaches to Stop Hypertension (DASH) (high intake of fruits, vegetables, low-fat dairy, grains, nuts, legumes, white meat, fish, and seasonings) are the best dietary model to reduce the risk of cardiovascular heart diseases and increase life expectancy in the 80 and above age range (Bhupathiraju, S., and Katherine, T., 2011; Lasheras, C et al., 2000; Estruch, R et al., 2013). Studies remain consistent on the conclusion that a healthy and good quality diet can prevent cardiovascular disease (Estruch, R et al., 2013; Lasheras, C et al., 2000). Researchers have only concluded that the Mediterranean diet pattern can prevent cardiovascular disease in the age group of 80 years and above (Lasheras, C et al., 2000). In addition, a healthy diet affects a person’s oral health and can prevent tooth decay with proper hygiene and lifestyle choices (Moynihan, Paula, 2005). Due to the importance of healthy diets in preventing cardiovascular disease and the likely interaction with oral and individual variables, it is essential to conduct this research and include the diet-related variables in the fixed-effects models. I used food insecurity %, the number of grocery stores, supercenters, convenience stores, and farmer’s markets to measure nutrition/ diet.

**Junk food consumption domain:**

In addition to a healthy diet, the consumption of junk food is a significant factor to study because of its influence and relation with oral health and heart disease development. Junk food is unhealthy food that includes fats, sugar, high sodium, low dietary fiber, refined carbohydrates, high levels of calories, highly processed meat, and fast food (Bhupathiraju, S., and Katherine, T., 2011). Researchers found a strong positive relationship between the risk of cardiovascular disease mortality and consuming a junk food diet (Temple, Norman J., 2018). Excessive consumption of processed carbohydrates and saturated fats can lead to obesity, a significant risk factor for heart failure and other heart diseases. Although studies found relationships between heart disease risk and junk food diets, researchers have not determined if changes in unhealthy diets may reverse the development of heart disease. For the predictive model, I used the number of fast-food and full-service restaurants to measure junk-food consumption.

**Individual factors domain:**

I used age levels and gender levels to measure individual factors in the fixed effects models to analyze heart disease mortality. Individual factors such as age and gender are non-mutable and vital risk factors for the risk of cardiovascular heart disease mortality. There are ongoing studies of heart disease because previous studies considered heart disease to mainly affect males. Still, heart disease is a significant cause of death for women in the United States. Studies have found a significant imbalance of heart disease mortality and education on heart disease between male and female patients (Hilleary, R. et al., 2019; Antelmi, I. et al., 2004; Stacy W., Nanette W., 2016). Regarding age variables, studies found that cardiovascular heart disease mortality was more common in elderly age groups (65 years and older), and the risk will continue to increase with age (Marelli, A. et al., 2007). Some studies have found patterns regarding age and other risk factors include diet, nutrition, and lifestyle (Moynihan, Paula, 2005). As people’s age increases, they are advised to adopt healthier diets and lifestyle choices to lower the risk of heart disease. But the interaction of a healthy diet and oral health as age progress has not been clearly defined yet (Batty, D et al., 2018; Estruch, R et al., 2013).

**Methods:**

**Dataset:**

I collected datasets from the Centers of Disease Prevention, Center Wonder (CDC Wonder, 2020), and the U.S. Department of Agriculture (USDA). The CDC Wonder dataset included years from 1999 to 2019, state, state abbreviation, deaths, crude death rate, age, and gender (CDC Wonder, 2020). The CDC datasets included variables for adults who had tooth removal due to tooth decay or gum disease and adults who visited dental clinics in the past year, states, and years (“Oral Health Data: Explore by Topic.”, 2020). The food environment dataset included variables for numbers of grocery stores, supercenter stores, convenience stores, fast-food restaurants, full-service restaurants, farmer’s markets, and percent household food insecurity per state. The food environment dataset included 2007, 2008, 2009, 2011, 2012, 2014, and 2016. Lastly, the USDA dataset included columns for states and counties (USDA, 2020).

**Data Wrangling:**

The data wrangling process consisted of cleaning the datasets and imputing missing data in R to create the predictive models and visuals. After collecting the datasets, I cleaned and manipulated the datasets in R to fit into the fixed-effects models. For the oral health and food environment data sets (CDC, 2019; USDA, 2020), I calculated new values for the variables by state by grouping by year and state. For the oral health data sets, I created new columns for state names by converting the state abbreviation column into the full state names. The data sets for oral health and food environment were downloaded as separate files for each year, I combined all the data sets together in decreasing year order after creating the new columns.

I used spline interpolation to fill in the missing data values for the risk factors. Spline interpolation is one of the methods under linear interpolation. The method was able to fill in missing values for certain years and takes into account both the state and year observations. Spline interpolation is a univariate time series imputation that can estimate missing data that belong to time-series datasets, which is an excellent fit for my dataset since the rows range from 1999 to 2019. The imputation of missing data can be predicted within the range of the original dataset, and the imputed data are observed over time. The imputeTS package in R has a function for linear interpolation imputation (Moritz, S. and Thomas B., 2017).

**Data Assumptions:**

The number of public datasets and variables in the datasets affected the construction of the model and the analysis. The heart disease mortality (crude rate per 100,00) dataset from CDC Wonder was the primary dataset and already adjusted heart disease mortality based on age groups and gender. Due to the age and gender variables, the heart disease deaths did not consider other factors (CDC Wonder, 2020). Regarding the oral health datasets, several people may not have reported their teeth had been removed or did not visit their dental clinic (CDC, 2019).

**Ethical Considerations:**

All datasets gathered from the CDC, CDC Wonder, and USDA have made their datasets free and publicly available to use. According to the CDC’s “Use of Agency Materials”, their information is in the public domain and can be used without the need for copyright permission (CDC). According to the CDC Wonder’s “Underlying causes of death, 1999 to 2919” dataset, the organizations restrict users to only use the data for health statistical reporting and analysis. The USDA has listed their datasets as public access level (USDA)

**Statistical Model:**

I constructed fixed-effect models to determine which risk factor variable best predicts heart disease mortality in the US from 1999 to 2019. Fixed effects models can predict change in variables within time-series data, evaluate interactions between variables and factor levels, and are common models in studies. A fixed-effect model assumes that the true effect of the variables will be identical for all studies, and assumes that the levels of categorical variables are independent of each other (Midway, S., 2021). The effect model can estimate the effects of categorical and numerical variables: age, gender, oral health, and food environment variables. The predictive models can determine the significant predictors and answer how each risk factor affects heart disease mortality (Midway, S., 2021). The dependent variable was the heart disease mortality crude rate (per 100,000). The independent variables that were fitted into the model were the year, state, age, gender, teeth removal, dentist visit, food insecurity %, and the number of grocery stores, convenience stores, supercenter stores, fast food restaurants, full-service restaurants, and farmer’s market per state.

**Results:**

**Exploratory data analysis**

I completed exploratory data analysis to identify outliers and variables with skewed values and observe patterns of the dataset before creating the fixed-effect models and data visuals. In the appendix section (Figure 1), I created a boxplot of heart disease mortality across years to check the distribution and skewness of the data. The box plot showed that all heart disease deaths for each year were higher than the median value, which meant a transformation of the data’s skewness was needed, and the data had a common variance. Then I created a scatter plot (Figure 2 in appendix) to study the relationship between heart disease mortality and the predictor variables. I decided to use the relation between teeth removal and heart disease death as the template for filtering outliers because all other predictors’ relations with heart disease were similar to it and don’t have strong correlations. Figure 2 showed two large clusters of data points, the cause of two clusters was due to the age range “85+” so I filtered out that age range. Figure 3 showed the scatter plot of teeth removal vs. heart disease mortality with the age filter and I realized there are a few outlier states. I filtered out the outlier states by creating scatter plots of teeth removal vs. heart disease death for each year (Figures 4 to 9 in the appendix). Figure 10 in the appendix showed the scatter plot of teeth removal and heart disease death with the age and state filters. Afterward, I created histograms of heart disease mortality and all the predictors (Figures 11 to 19 in the appendix) to observe whether the data is skewed. The graphs would help me know which variables needed to be log-transformed so all variables would have a normal distribution and not affect the predictive model. The variables that were log-transformed were: grocery stores, convenience stores, and farmer’s markets. I added the log-transformed variables to the dataset in order to create the fixed-effect models.

**Statistical Analysis:**

Fixed-effect models were constructed to determine the significant predictor variables for heart disease mortality crude rate (per 100,000). An interactive map, trend graphs, and scatter plots were created to observe the patterns of the variables. Six fixed effects models were created: the first model included all independent variables, the second model only included variables after the VIF test, and the rest of the models had filtered age group because the age variable caused multiple structures of the data. The models for the age group only contained the numerical variables to avoid overfitting and causing different structures of the data.

Table 1 showed the independent and dependent variables that were fit in the fixed-effects model. The VIF test was used to validate the fixed-effects model by checking for multicollinearity in the data. Since the fixed-effects model contains multiple variables and factors, it can cause highly correlated variables. Multicollinearity is to be avoided because it could cause unreliable results and larger confidence intervals (Salmerón-Gómez, R., Catalina García-García, and José García-Pérez, 2021). Additionally, this project aimed to understand the relation between the independent and dependent variables as best as possible. Therefore, the VIF test will be the best method to measure the multicollinearity value between variables. A high VIF score indicates high collinearity between the independent variable and the predictor variables (Daoud, Jamal I, 2017).

Table 2 showed the results of the VIF multicollinearity test for the fixed-effect model. The table lists all the predictor variables and their corresponding VIF scores. All predictors with a VIF score of 5 or higher were removed from the updated fixed-effect models because the variable has a high correlation with other variables. The independent variables removed were year, age, grocery stores, convenience stores, fast-food restaurants, and full-service restaurants.

Table 1 in the appendix showed the residuals of the updated fixed-effects model after the VIF test. The residual of the model was the difference between the observed data points of heart disease mortality and the predicted values for the dependent variable. The range of the residuals implied that a curve would not be fit for this data and there was not a consistent variance.

|  |  |  |
| --- | --- | --- |
| **Model** | **Dependent variable** | **Independent variables** |
| Fixed-effect model | Heart disease mortality crude rate (per 100,000) | * State * Year * Age * Gender * Teeth removal due to cavities or gum disease * Dentist visits within the past year * Grocery stores per state * Convenience stores per state * Supercenters per state * Fast-food restaurants per state * Full-service restaurants per state * Food insecurity per state * Farmer’s market per state |
| Fixed-effect model after VIF test | Heart disease mortality crude rate (per 100,000) | * State * Gender * Teeth removal due to cavities or gum disease * Dentist visits within the past year * Supercenters per state * Food insecurity per state * Farmer’s market per state |
| Fixed-effect model with filtered age groups | Heart disease mortality crude rate (per 100,000) | * Teeth removal due to cavities or gum disease * Dentist visits within the past year * Supercenters per state * Food insecurity per state * Farmer’s market per state |

**Table 1:** Dependent and independent variables that were fit into the fixed-effect models.

|  |  |
| --- | --- |
| **Variable** | **GVIF** |
| Year | 30.83 |
| State | 4.32 |
| Age | 19.15 |
| Gender | 2.00 |
| Teeth removal | 1.94 |
| Dental visit | 2.08 |
| Grocery stores | 12.77 |
| Supercenters | 2.91 |
| Convenience stores | 8.53 |
| Fast Food restaurants | 14.58 |
| Full-service restaurants | 20.78 |
| Food insecurity | 2.21 |
| Farmer’s market | 1.88 |

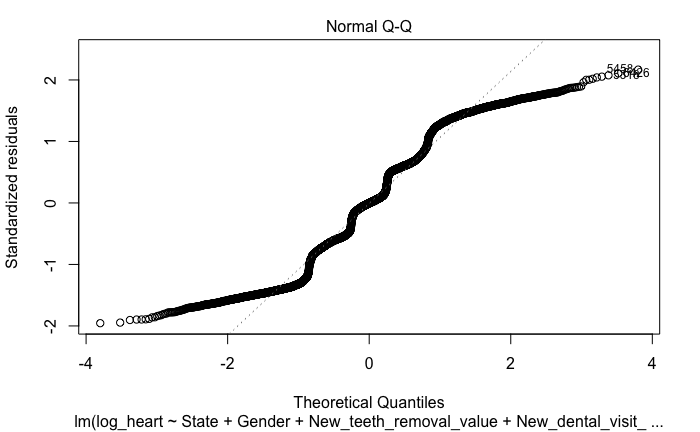
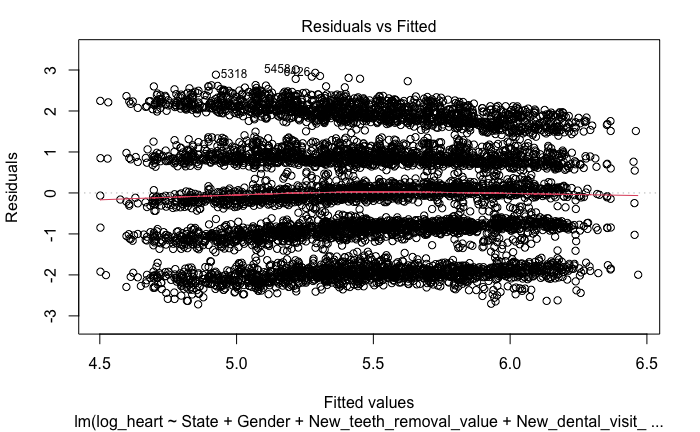
**Table 2:** VIF test for validation of the fixed-effect model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficients** | **Estimate** | **Std. Error** | **t value** | **p-value** | **Coefficients** | **Estimate** | **Std. Error** | **t value** | **p-value** |
| Intercept | 5.12 | 4.02 | 12.74 | < 2e-16 \*\*\* | State Oklahoma | -5.77 | 1.36 | -0.42 | 0.67 |
| State Arizona | -5.96 | 1.36 | -4.35 | 1.23e-05\*\*\* | State Oregon | -7.26 | 1.37 | -5.30 | 1.19e-07 \*\*\* |
| State Arkansas | -6.92 | 1.38 | -0.51 | 0.61 | State Pennsylvania | -4.79 | 1.38 | -3.46 | 0.00 \*\*\* |
| State California | -6.03 | 1.38 | -4.35 | 1.33e-05 \*\*\* | State South Carolina | -2.00 | 1.36 | -1.46 | 0.14 |
| State Colorado | -7.99 | 1.40 | -5.78 | 7.63e-009 \*\*\* | State Tennessee | -1.29 | 1.36 | -0.94 | 0.34 |
| State Connecticut | -7.48 | 1.37 | -5.32 | 1.04e-07 \*\*\* | State Texas | -3.21 | 1.38 | -2.32 | 0.02 \* |
| State Florida | -5.37 | 1.36 | -3.91 | 9.12e-05 \*\*\* | State Virginia | -5.19 | 1.39 | -3.71 | 0.00 \*\*\* |
| State Georgia | -2.03 | 1.37 | -1.48 | 0.13 | State Washington | -7.06 | 1.37 | -5.12 | 2.99e-07 \*\*\* |
| State Indiana | -3.43 | 1.38 | -2.49 | 0.012520 \* | State Wisconsin | -6.23 | 1.39 | -4.54 | 5.54e-06 \*\*\* |
| State Iowa | -5.44 | 1.37 | -3.9 | 8.46e-05 \*\*\* | State New Jersey | -6.58 | 1.38 | 4.74 | 2.17e-06 \*\*\* |
| State Kansas | -5.07 | 1.36 | -3.69 | 0.00 \*\*\* | State New York | -5.29 | 1.38 | -3.82 | 0.00 \*\*\* |
| State Kentucky | -1.56 | 1.38 | -1.14 | 0.25 | State North Carolina | -3.26 | 1.36 | -2.39 | 0.02 \* |
| State Maryland | -4.46 | 1.40 | -3.22 | 0.00 \*\* | State Ohio | -3.15 | 1.37 | -2.30 | 0.02 \* |
| State Massachusetts | -8.26 | 1.38 | -5.89 | 3.84e-09 \*\*\* | Gender Male | 6.25 | 3.35 | 1.86 | < 2e-16 \*\*\* |
| State Michigan | -3.22 | 1.38 | -2.33 | 0.019 \* | Number\_Teeth\_  removal | 1.05 | 3.23 | 3.26 | 0.001 \*\* |
| State Minnesota | -9.28 | 1.41 | -6.59 | 4.66e-11 \*\*\* | Number\_dental\_  visits | 7.17 | 4.02 | 1.78 | 0.07 \* |
| State Missouri | -2.34 | 1.38 | -1.72 | 0.08 \* | Number\_  supercenters | -5.75 | 2.34 | -0.24 | 0.80 |
| State Montana | -2.20 | 1.38 | -1.59 | 0.11 | Foodinsecurity\_  percentage | -1.84 | 7.29 | -2.52 | 0.01 \* |
| State Nevada | -2.98 | 1.38 | -2.16 | 0.03 \* | log\_farmers\_market | -5.107 | 2.65 | -1.92 | 0.05 \* |
| State New Hampshire | -4.00 | 1.41 | -2.82 | 0.004 \*\* |  |  |  |  |  |

**Table 3:** Shows all the predictor’s and dependent variables’ corresponding slope estimates, standard error, t- value, and p-value.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Residual standard error** | **Multiple R-squared** | **Adjusted R-squared** | **F-statistic** | **p-value** |
| Fixed-effect | 1.39 on 6874 degrees of freedom | 0.07 | 0.07 | 15.04 on 38 and 6874 DF | < 2.2e-16 |

**Table 4:** Shows the resulting measurements for the significance of the fixed-effect model



**Figure 1:** Plots show residuals and normality of the fixed-effects model in table 3

Table 3 showed the coefficients and corresponding estimates, standard error, t-value, and p-values that were fit into the fixed-effects model after the VIF test. The estimated value for the intercept was the slope of the dependent variable, which was the expected average of heart disease mortality. Significant predictors had p-values less than 0.05. The estimates of the coefficients are their average effect on heart disease mortality, in other words, the estimates are the slope average in their groups. The standard error values are the average distance of the actual data points from the regression line. The t values represent the accuracy of the coefficient’s estimates. The significant risk factors for heart disease deaths were teeth removal, dentist visits within the last year, food insecurity % per state, and the number of farmer’s markets per state. The majority of state levels and the male gender were significant predictors. The number of dentist visits within the past year and the number of teeth removed per state had positive effects on heart disease mortality. This implied that a rise in heart disease death by one unit can be in relation to an increase in the number of dentist visits and the number of teeth removed by their corresponding estimate values. Food insecurity percentage and number of farmer’s markets per state had negative effects on the dependent variable. Which imply that a rise in heart disease death by one unit can be in relation to a decrease in food insecurity % and the number of farmer’s market by their corresponding estimate values. For the categorical variables, the reference level for State was Alabama and the reference level for gender was female. All significant states had negative effects on heart disease mortality in reference to the effects of Alabama. Males had a positive effect on heart disease mortality when compared to females.

Table 4 showed the fixed-effects model’s (Table 3) residual standard error, multiple R-squared, adjusted R-squared, F-statistic, and p-value. Since the p-value of the model is less than 0.05, the model is significant. The adjusted R-squared equals 0.07 which meant that the predictor variables explained 7% of the variance in heart disease mortality in the model.

Figure 1 showed the residuals and normal distribution of the fixed-effects model in table 3 and the plots are used for model validation. The graph on the left showed five groups of residuals and the third group of residuals were clustered around the red line with some outliers. The multiple groups of residuals were caused by the age predictor because the five different levels within that predictor caused multiple structures of the data. This meant separate models would need to be made for each age group. The residual group that gathered on the red line implies the model didn’t have heteroscedasticity and the dependent variable has large values. The outliers may be caused by data points with high variation from the predicted values. The graph on the right shows that the residuals have a rather wavy line and did not directly fall on the dotted line. This implies the age levels influenced the model’s normality and would need to filter the age groups in order for the other predictor fitted into the model to have a normal distribution.

Tables five through nine showed the fixed effects model and significance for each age group: 35-44, 45-54, 55-64, 65-74, and 75-84. Figures 20 to 24 in the appendix show the residual and normality plots of the models. The 35-44 and 55-64 age groups had the number of dentist visits, the number of teeth removed, and the number of farmer’s markets per state to be significant predictors for heart disease death. The 45-54 age group had the same significant predictors as in the table 3 model. The 65-74 and 75-84 age groups had the number of teeth removed, food insecurity %, and the number of farmer’s markets as significant predictors for heart disease death. All age group models had a significant p-value and their adjusted R-squared values in order are the following: 0.027, 0.064, 0.06, 0.127, and 0.214. When filtering by age group, the fixed-effects model explained the variation of heart disease death within the range of 2.7% - 21.4%. Within the age group models, the majority has the number of dentist visits, food insecurity %, and the number of farmer’s market variables to have a negative effect on heart disease death. This implies a rise in heart disease death by one unit can be related to a decrease in those variables by their corresponding estimate values. The age group models have the number of teeth removed variable to have a positive effect on heart disease deaths. This implies a rise in heart disease death by one unit was related to an increase in the number of teeth removed by its estimated value. Older age groups (65-74 and 74-84) have higher adjusted R-squared values, which meant older age can explain more of the variation in heart disease deaths than younger age groups in relation to the predictors fitted into the models. Furthermore, Figures 20 to 24 in the appendix show the residual and normality plots of the age group fixed-effects models. The residual plots on the left of Figures 20-24 show residuals clustered around the red dotted line with a few outliers. The residual pattern implies the models did not have heteroscedasticity and the outliers may be caused by variation in the dependent variable. The normality plots on the right of Figure 20-24 show the residuals to lie straight on the dotted line with some outliers. This means the data fit into the age filter models have normal distributions and the outliers may be caused by the range of values in the variables.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **log\_heart** | | | |
| *Predictors* | *Estimates* | *Std. error* | *t-value* | *p* |
| (Intercept) | 4.47 | -0.00 | 10.20 | **<0.001** |
| New teeth removal value | 0.01 | 0.08 | 2.71 | **0.007** |
| New dental visit value | -0.01 | -0.09 | -3.05 | **0.002** |
| New supercenter value | -0.00 | -0.04 | -1.23 | 0.220 |
| New food insecurity value | 0.01 | 0.03 | 1.07 | 0.287 |
| log farmers | -0.12 | -0.11 | -3.82 | **<0.001** |
| Observations | 1369 | | | |
| R2 / R2 adjusted | 0.031 / 0.027 | | | |

**Table 5:** Fixed effects model result of age group 35-44

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **log\_heart** | | | |
| *Predictors* | *Estimates* | *Std. error* | *t-value* | *p* |
| (Intercept) | 5.38 | 0.00 | 18.31 | **<0.001** |
| New teeth removal value | 1.18 | 0.12 | 4.56 | **<0.001** |
| New dental visit value | -1.80 | -0.17 | -6.03 | **<0.001** |
| New supercenter value | 5.76 | 0.00 | 0.03 | 0.975 |
| New food insecurity value | 1.38 | 0.07 | 2.50 | **0.012** |
| log farmers | -7.46 | -0.10 | -3.58 | **<0.001** |
| Observations | 1386 | | | |
| R2 / R2 adjusted | 0.067 / 0.064 | | | |

**Table 6:**  Fixed effects model result of age group 45-54

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **log\_heart** | | | |
| *Predictors* | *Estimates* | *std.error* | *t-value* | *p* |
| (Intercept) | 6.14 | -0.00 | 21.77 | **<0.001** |
| New teeth removal value | 0.01 | 0.15 | 5.63 | **<0.001** |
| New dental visit value | -0.01 | -0.14 | -5.19 | **<0.001** |
| New supercenter value | 0.00 | 0.02 | 0.85 | 0.394 |
| New food insecurity value | -0.00 | -0.01 | -0.39 | 0.699 |
| log farmers | -0.07 | -0.10 | -3.61 | **<0.001** |
| Observations | 1386 | | | |
| R2 / R2adjusted | 0.064 / 0.060 | | | |

**Table 7**: Fixed effects model result of age group 55-64

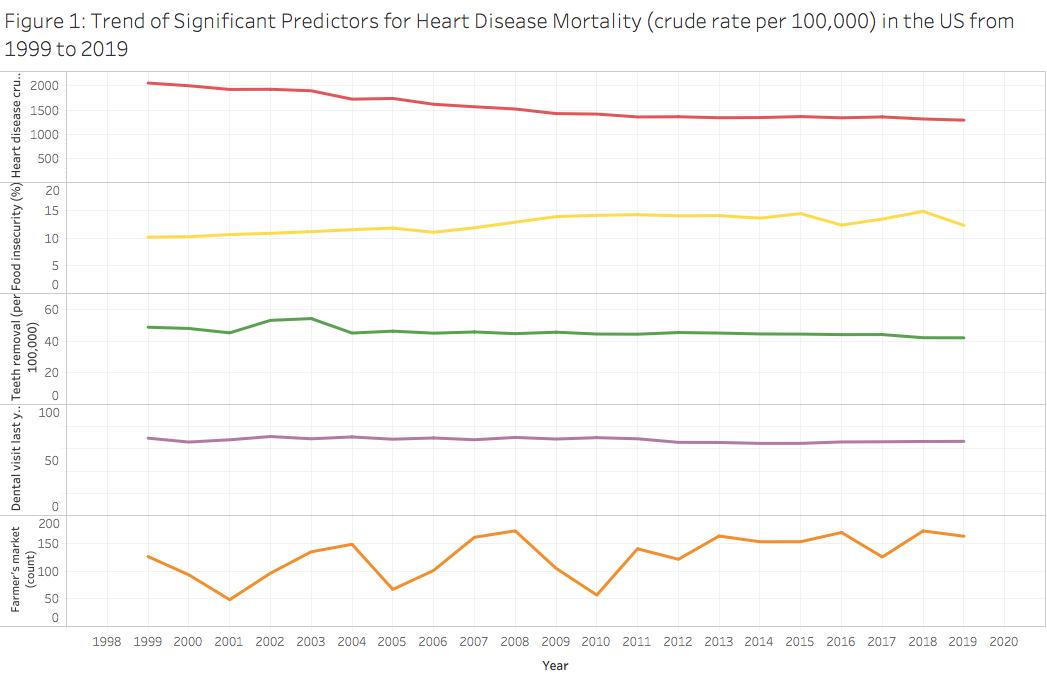
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **log\_heart** | | | |
| *Predictors* | *Estimates* | *Std. error* | *t-value* | *p* |
| (Intercept) | 6.42 | -0.00 | 28.20 | **<0.001** |
| New teeth removal value | 0.02 | 0.25 | 9.60 | **<0.001** |
| New dental visit value | -0.00 | -0.04 | -1.47 | 0.141 |
| New supercenter value | -0.00 | -0.00 | -0.13 | 0.899 |
| New food insecurity value | -0.02 | -0.15 | -5.58 | **<0.001** |
| log farmers | -0.08 | -0.14 | -5.34 | **<0.001** |
| Observations | 1386 | | | |
| R2 / R2adjusted | 0.130 / 0.127 | | | |

**Table 8:** Fixed effects model result of age group 65-74

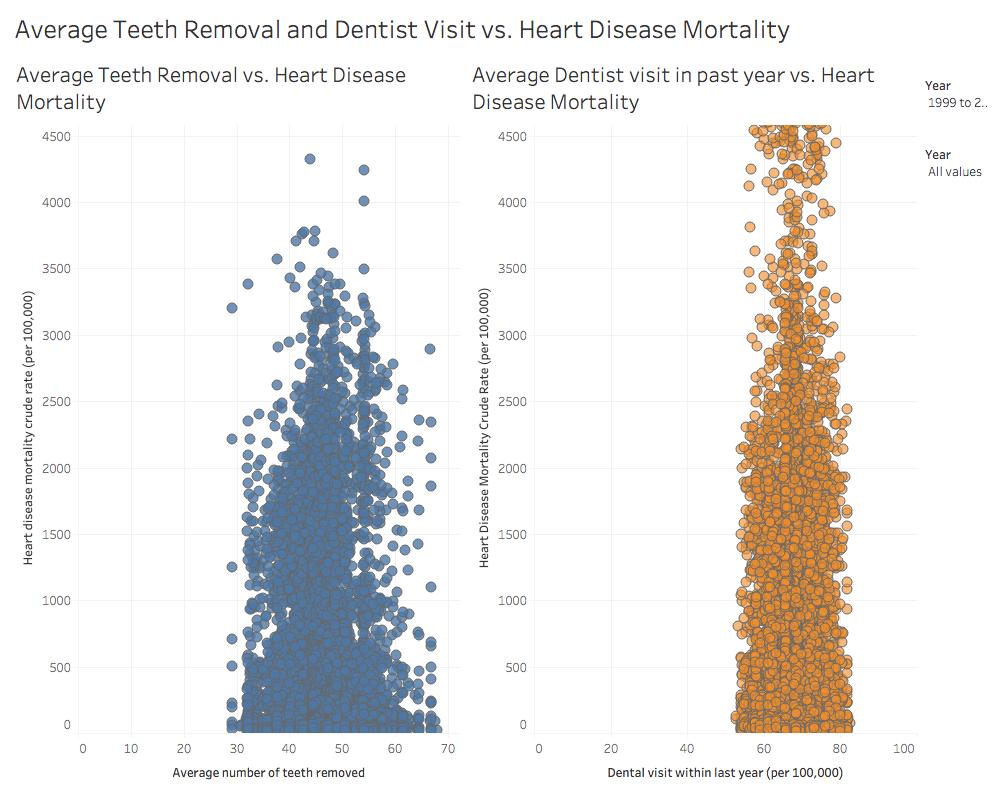
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **log\_heart** | | | |
| *Predictors* | *Estimates* | *std.error* | *t-value* | *p* |
| (Intercept) | 7.20 | 0.00 | 46.14 | **<0.001** |
| New teeth removal value | 0.02 | 0.31 | 12.55 | **<0.001** |
| New dental visit value | 0.00 | 0.05 | 1.85 | 0.064 |
| New supercenter value | -0.00 | -0.00 | -0.13 | 0.899 |
| New food insecurity value | -0.03 | -0.23 | -8.97 | **<0.001** |
| log farmers | -0.08 | -0.18 | -6.98 | **<0.001** |
| Observations | 1386 | | | |
| R2 / R2adjusted | 0.217 / 0.214 | | | |

**Table 9**: Fixed effects model result of age group 75-84

I created trend graphs (Figure 2) to observe changes in heart disease mortality and the significant predictors across all fixed effects models. The red line represented the heart disease mortality trend, the yellow line represented the food insecurity % trend, the green line represented the number of teeth removed trend, the purple line represented the number of dentist visits trend, and the orange line represented the number of farmer’s market trend. Figure 2 showed that the average number of heart disease mortality crude rate (per 100,000) decreased from 1999 to 2019. The food insecurity trend remained constant across the years along with a steady increase since 2006. Both teeth removal and dentist visits had a rather linear trend across the years. The number of farmer’s market trends is inconsistent with a sharp increase or decrease in numbers in certain years. Thus, Figure 2 shows no apparent similar pattern between heart disease death and the significant patterns from 1999 to 2019.

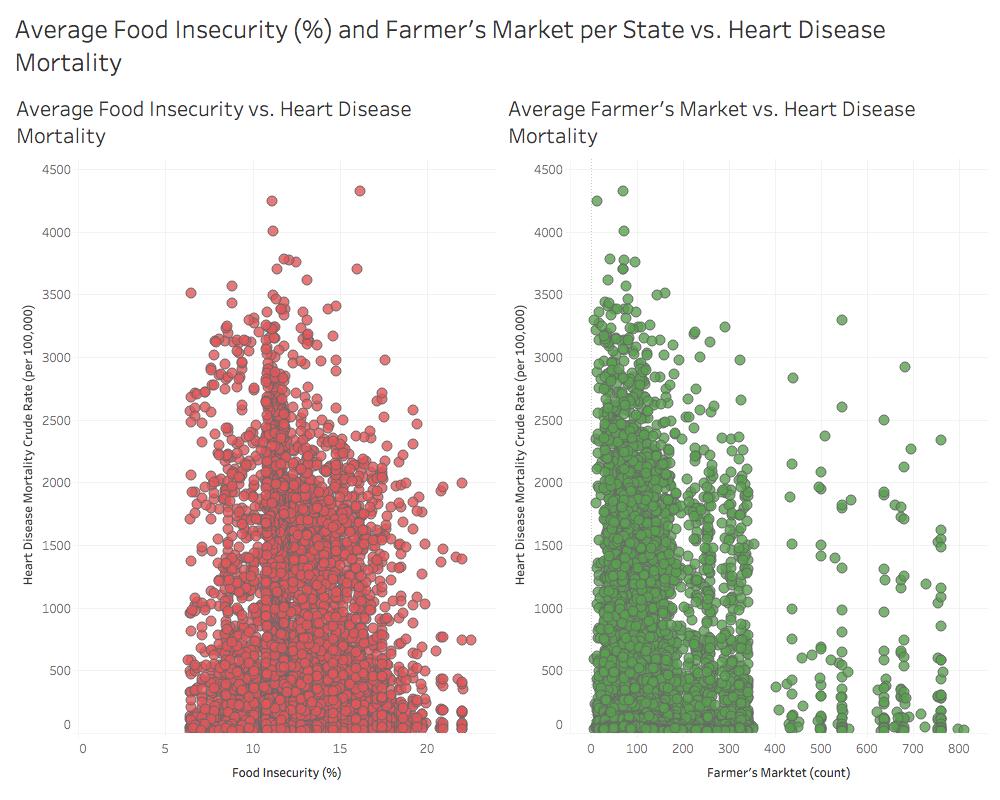
****

**Figure 2:** Trend graph showing the change in the average heart disease death, teeth removal, dentist visit, food insecurity %, and the number of farmer’s markets across years

****

**Figure 3:** Scatter plots of number of teeth removal and dentist visits and heart disease mortality.

I created scatter plots to determine if the significant predictors had correlations with heart disease mortality since the trends of the predictors in Figure 2 were rather linear. Figure 3 showed that there were weak correlations between the average number of teeth removal and the average number of dentist visits with average heart disease mortality. The data values displayed in Figure 3 included all age groups.

****

**Figure 4:** Scatter plots show the average food insecurity % and farmer’s market in relation to heart disease mortality

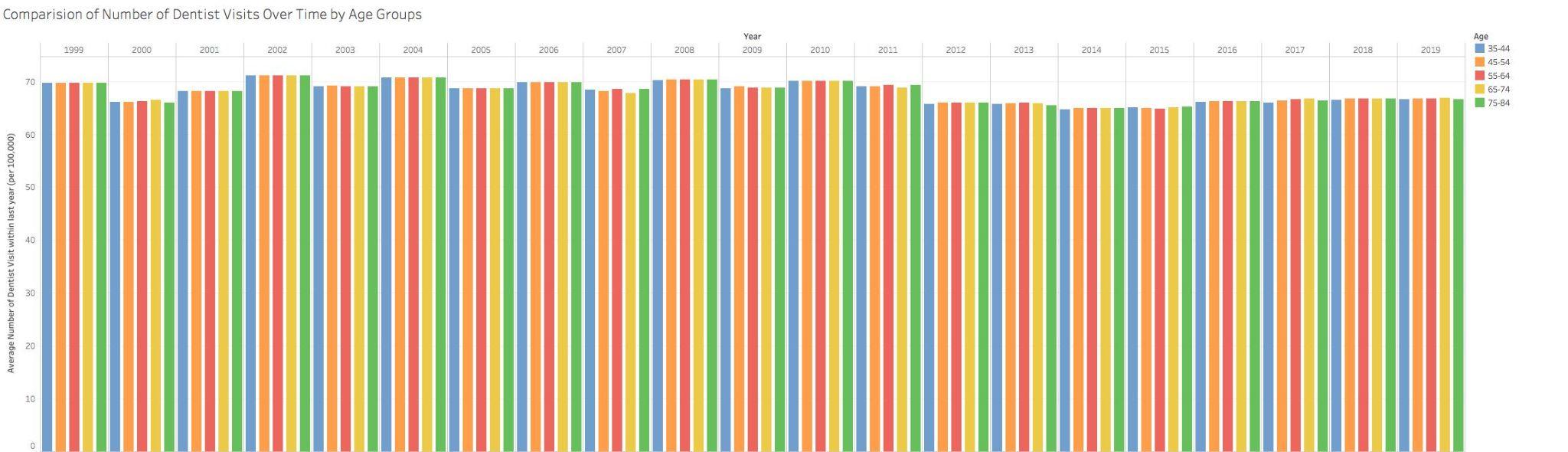
Figure 4 showed that there were weak correlations between average food insecurity % and the number of farmers’ markets with average heart disease death. The data values displayed in Figure 4 included all age groups.

The grouped bar graphs (Figures 5 to 9) show the change in heart disease deaths and the significant predictors by age groups. Since separate fixed-effects models were created for each age group, it was best to observe if the trends of the variables differed depending on age. In Figure 5, the decrease in heart disease deaths across years resembles the pattern in Figure 2 and older age groups, 65-74 and 75-84, had higher numbers of heart disease deaths. Thus, older age groups are more likely to die of heart disease. In Figures 6 to 9, the trends of the significant predictors resemble the linear trend in figure 2, and the trends did not have significant differences depending on age groups.

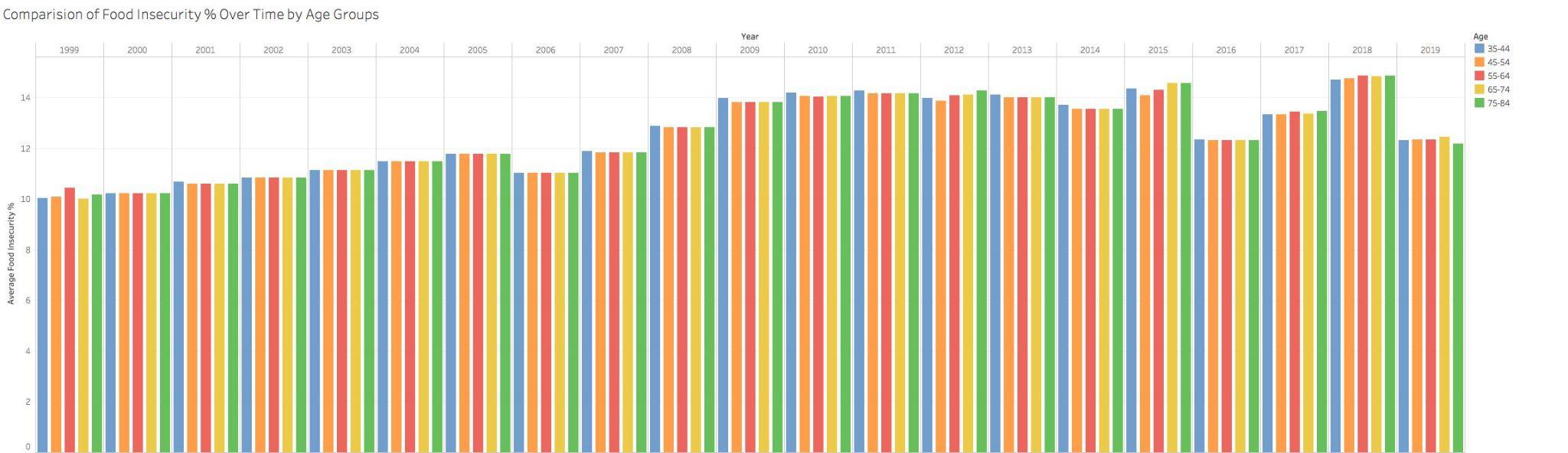
**Figure 5:** Change in heart disease mortality over years by age groups



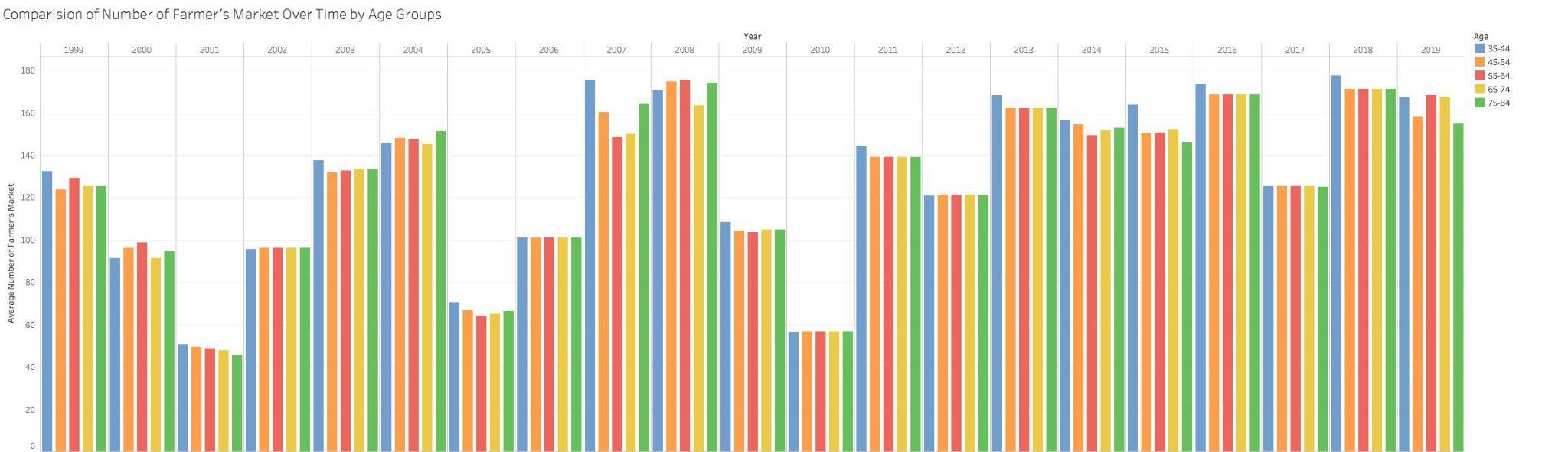
**Figure 6:** Change in teeth removed variable over years by age group



**Figure 7:** Change in dentist visit variable over years by age group

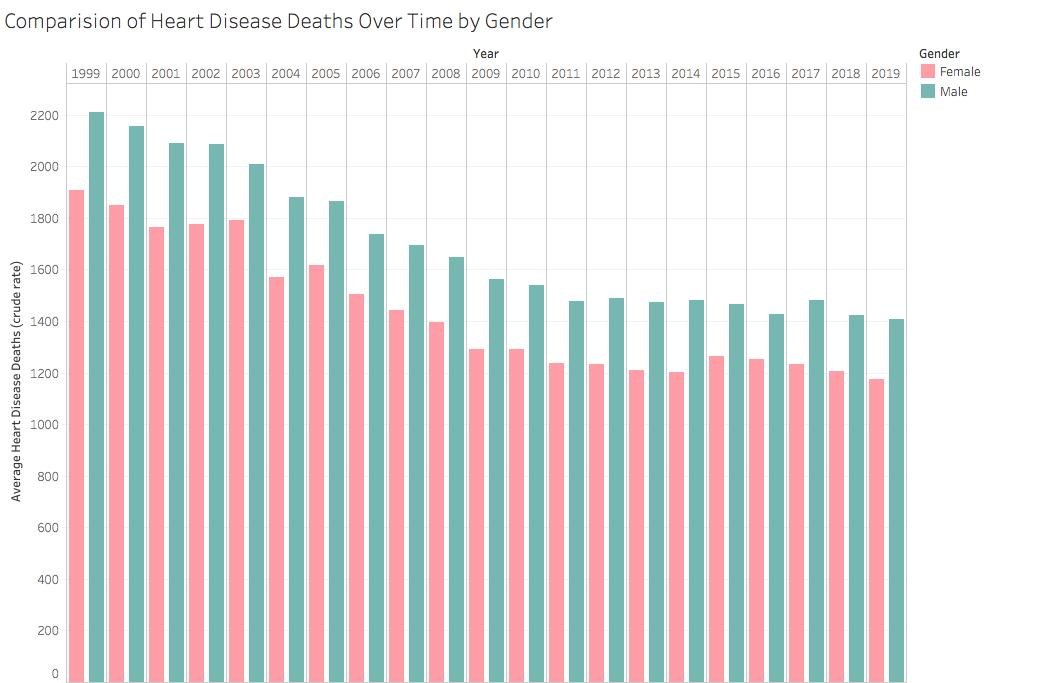


**Figure 8:** Change in food insecurity % variable over years by age group

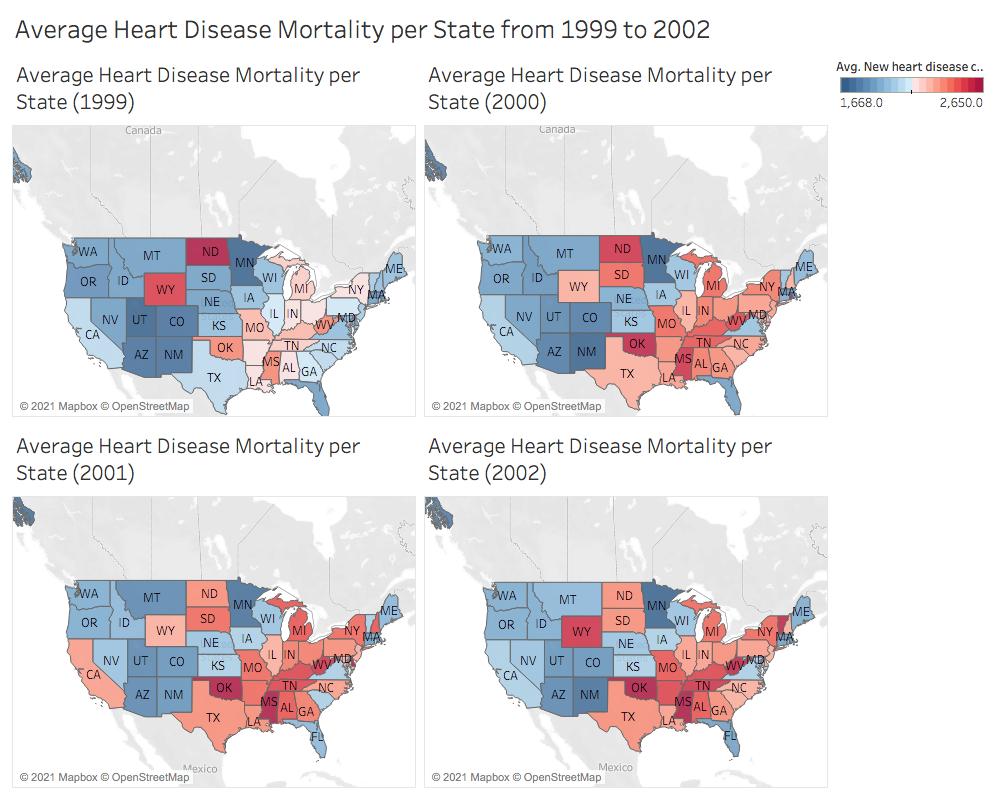


**Figure 9:** Change in number of farmer’s market variable over years by age group

I created a grouped bar graph (Figure 10) to determine if heart disease mortality differed by gender from 1999 to 2019. All age groups (35 - 84) were included in the data values of figure 10. The blue bars represented males and the pink bars represented females. Figure 10 showed that males had significantly more heart disease deaths compared to females in the US from 1999 to 2019. The decreasing trend of heart disease deaths resembled the trend in figure 2, and for each year males continued to have higher numbers of reported heart disease deaths than females.



**Figure 10:** Change in heart disease mortality



**Figure 11:** US map of average heart disease mortality by color (1999 to 2002)

Figure 11 is a snapshot example of my interactive map dashboard on Tableau titled “Map dashboard”. The interactive map dashboard showed the US map and included values for the state name, heart disease mortality (per 100,00), number of teeth removal, number of dentist visits, food insecurity %, and number of farmer’s markets per state and per year. On the side of the map is a year filter where users can adjust the range of years they want to observe. Blue-colored values indicated a lower heart disease death crude rate and red-colored values indicated a higher heart disease death crude rate. The analysis of the map is drawn from 5-year ranges due to the variation in heart disease mortality crude rate and different population sizes of each state. From 1999-2003: the majority of the eastern and southern states had an increase in heart disease mortality. Also, Mideastern states such as Missouri and West Virginia were more consistent in having high numbers of heart disease deaths. From 2004-2008: Southern and some Midwestern states including Missouri, Oklahoma, Tennessee, and Mississippi had a consistent high range of heart disease mortality. From 2009-2013: a majority of Southern and Midwestern states in the previous 5-year range maintained high heart disease mortality value. And there tended to be lower ranges of heart disease deaths. From 2014-2019: mainly the Southern states including Mississippi and Alabama had consistent high heart disease mortality rates. Overall, the majority of states had lower heart disease death values in that year range.

**Discussion:**

This research project contributed to research in heart disease risk factors by exploring how some health factors that are related to heart disease mortality. Some of the results are supported by previous studies on heart disease whereas other results contradicted the findings of previous studies. In the study conducted by Batty, (Batty, D et al., 2018) tooth cavities have a relation with heart disease and found that people with tooth cavities have an increased risk of heart disease, which supports an aspect of my fixed-effects model result. As shown in the fixed effects model in table 3, the number of teeth removed due to tooth cavities of gum disease variable had a positive effect on heart disease. Which meant a rise in heart disease deaths can be related to an increase in the number of teeth removed per state. The number of teeth removed due to cavities or gum disease may be a measure of the severity of people’s poor oral health. Higher numbers of teeth removed can be used as an indication of poor oral health status and may have an increased risk of dying from heart disease (Batty, D et al., 2018). On the other hand, the study conducted by Sanchez (Sanchez, Paula, et al., 2017) did not support the model’s finding on the dentist visit variable. The fixed-effects model in table 3 found that the number of dentists visit had a positive effect on heart disease deaths. Which meant a rise in heart disease deaths can be related to an increase in the number of dentist visits. However, Sanchez’s study found that people who left their dental problems untreated (not going to the dentist to have a dentist resolve the dental issue) had more association with heart disease problems (Sanchez, Paula, et al., 2017).

In the fixed-effects model, the number of farmer’s markets variable was a significant predictor used to measure access to maintain a healthy diet. The models concluded that the farmer’s market variable had a negative effect on heart disease mortality. This meant a rise in heart disease deaths is related to a decrease in the number of heart disease deaths. The effects of the farmer’s market variable were supported by Heidemann’s study. According to Heidemann’s research (Heidemann, C et al., 2008), healthy diets composed of a rich intake of fruits and vegetables can decrease the risk of heart disease. Farmer’s markets are a great resource to get fresh, organic, and nutritional products. When there is less access to healthy food options, people are not able to maintain a healthy diet and get the nutrition needed to live long and healthy lives (Heidemann, C et al., 2008). Lastly, other studies found that higher consumption of fruits and vegetables leads to a lower risk of cardiovascular heart disease deaths ( Lasheras, C et al., 2000).

The food insecurity % per state variable was also used to measure the healthy diets and nutrition intake of people in the US. The fixed-effects model determined that the food insecurity variable had a negative effect on heart disease mortality. This meant a rise in heart disease deaths is related to a decrease in food insecurity. The effects of the food insecurity variable contradict the results of other studies. According to Mendy’s research (Mendy, Vincent L, et al., 2018), food insecurity was associated with high blood pressure, diabetes, obesity, physical inactivity, and smoking, which were all risk factors for heart disease. Food insecurity is defined as households having limited access to nutritional food sources in certain economic and social conditions (USDA). The hypothesis of a decrease in food security is associated with higher heart disease risk remained inconclusive in studies (Liu, Yibin, and Heather A Eicher-Miller, 2021).

The fixed-effects models and graphs determined that males had higher numbers of heart disease deaths compared to females. Previous studies found heart disease to have more effect on males in recent years (Vaughan AS, Schieb L, Casper M., 2020). More reported deaths in males that were caused by cardiovascular heart disease may be caused by the classification of heart disease on death certificates since the symptoms, risk factors, and behavior of heart disease differ by gender (Vaughan AS, Schieb L, Casper M., 2020). Some studies found that the male death rate caused by heart disease may be associated with smoking and alcohol consumption behavior (Hilleary, R. et al., 2019). On the other hand, Hilleary’s study found that the majority of people who received treatment for heart disease were white, non-Hispanic/ Latino, and males were 0.86 times more likely to receive treatment than women (Hilleary, R. et al., 2019). In addition, women received less patient education on heart disease than males. Women received less counseling care that would help them improve their lifestyle and better health management. In addition, physicians were more likely to prescribe pain relief medications to male patients than female patients. Males receiving more cautious treatment can imply that women’s pain and symptoms are taken less seriously, underestimating cardiovascular heart disease diagnosis. The conclusion that females received less patient education than males still needs further evidence. Moreover, more research is still required to evaluate how the interaction of other risk factors, patient education, and sociodemographic variables could cause the imbalance of heart disease mortality between gender (Hilleary, R. et al., 2019; Antelmi, I. et al., 2004; Stacy W., Nanette W., 2016).

The graphs determined that older age groups (65 to 84) had higher numbers of heart disease deaths compared to younger age groups (35-64). The fixed-effects model with filtered age group found that the dentist visits variable had positive effects in the oldest age group (74-84) whereas the younger age groups had negative effects for those variables. This implies for younger people, fewer visits to the dentist to treat tooth problems can be related to increased heart risk. This result is also supported by other studies (Sanchez, Paula, et al., 2017). Furthermore, older age groups having more heart disease deaths compared to younger age groups can be explained by various factors. With old age, people’s organs start to become weaker and have less nutrition which can be associated with heart failure and heart disease-related risks (Marelli, A. et al., 2007). Studies have found certain patterns with age and other risk factors include diet, nutrition, and lifestyle. However, the interaction of a healthy diet and oral health as age progress has not been clearly defined yet (Batty, D et al., 2018; Estruch, R et al., 2013; Moynihan, Paula, 2005).

This research had several limitations which may explain the weak relations of the chosen risk factors with heart disease mortality. Due to the lack of available datasets and variables, the results of the model and graphs may not be the fittest to answer my research questions. I wanted to analyze time-series data sets of heart disease mortality across all US states, which meant the rows of the data sets had to be years and states. But the availability of that data set structure with specific variables and had public access was limited. I initially searched for datasets that included variables for different food consumptions, such as % consumption of sugar and vegetables, as measurements for healthy and unhealthy diet patterns. But I was not able to find an available dataset in the correct structure I wanted. Furthermore, the reliability of the fixed-effect model’s result is impacted by the variation in heart disease mortality. Because the heart disease mortality crude rate is based on the state’s yearly population. In addition, the weak correlation between the predictor variables and the dependent variable may have caused the model to overfit. However, the map and trend graphs correspond to the patterns of the variables, and the model validations and exploratory analysis concluded the model’s assumptions were met. Lastly, because it is a time-series dataset and all state levels are included in the dataset, it may help with the population variance.

Due to the limitations and results contradiction with previous studies, further research and improvement of the fixed effects models should be considered. Mixed-effect models and models with interaction terms could be used to further analyze heart disease mortality in relation to the predictor variables in time-series data sets. More data sets that contain variables related to healthy and unhealthy diet patterns could contribute to the model. According to previous research, other candidate variables that could be fit into future models are smoking habits, obesity, alcohol consumption, exercise/ lifestyle, different % consumption of food, and ethnicity (Batty, D et al., 2018; Estruch, R et al., 2013; Moynihan, Paula, 2005; Hilleary, R. et al., 2019).

**Conclusion:**

The fixed-effects models determined that the significant predictors for heart disease mortality were state levels, gender levels, number of teeth removed due to gum disease or cavities, number of dentist visits within the past year, food insecurity % per state, and number of farmer’s markets per state. Heart disease mortality decreased from 1999 to 2019 in the US. Males had higher numbers of heart disease deaths compared to females. Older age groups (65-74 and 75-84) had higher numbers of heart disease deaths (crude rate) compared to younger age groups (35 to 64). The significant risk factors did not differ depending on age groups. The majority of state levels, food insecurity %, and farmer’s markets had negative effects on heart disease mortality. The majority of teeth removal and dentist visits had positive effects on heart disease mortality. The model predicted that an increase in heart disease will result from higher numbers of teeth removed and dentist visits. Furthermore, the model predicted that a rise in heart disease deaths will result from a decrease in the number of farmers’ markets and food insecurity (%) per state. Though the models were significant, the model’s variables had weak relations and explained a low percentage of change in heart disease mortality across all age groups. According to the trend and bar graphs, the trend in significant risk factors was not similar to the trend of heart disease deaths from 1999 to 2019. The scatterplots showed that the significant predictors had weak correlations (rather linear trends) with heart disease mortality. This implied that the significant predictors found by the fixed-effects models only account for a low percentage that can explain the underlying relation between heart disease mortality and oral health. To improve the model, more diet pattern-related variables or fewer variables can be fit into the model. Due to the weak relations, the model may have been overfitted. The results of this project imply that there are other underlying factors that can be explored to study the association between heart disease deaths and health factors. Improvement of the model could become a beneficial tool in the long term for health professionals and researchers.

**References:**

Aune, Dagfinn et al. “Fruit and vegetable intake and the risk of cardiovascular disease, total

cancer and all-cause mortality-a systematic review and dose-response meta-analysis of

prospective studies.” *International journal of epidemiology* vol. 46,3 (2017): 1029-1056.

doi:10.1093/ije/dyw319

Antelmi, Ivana et al. “Influence of age, gender, body mass index, and functional capacity on

heart rate variability in a cohort of subjects without heart disease.” *The American journal*

*of cardiology* vol. 93,3 (2004): 381-5. doi:10.1016/j.amjcard.2003.09.065

Belle, Ashwin et al. “Big Data Analytics in Healthcare.” *BioMed research international* vol.

2015 (2015): 370194. doi:10.1155/2015/370194

Batty, G David et al. “Oral health and later coronary heart disease: Cohort study of one million

people.” *European journal of preventive cardiology* vol. 25,6 (2018): 598-605.

doi:10.1177/2047487318759112

Bhupathiraju, Shilpa N, and Katherine L Tucker. “Coronary heart disease prevention: nutrients,

foods, and dietary patterns.” *Clinica chimica acta; international journal of clinical*

*chemistry* vol. 412,17-18 (2011): 1493-514. doi:10.1016/j.cca.2011.04.038

“CDC Wonder.” *Centers for Disease Control and Prevention*, Centers for Disease Control

and Prevention, wonder.cdc.gov/.

Ciampi, A., et al. "Family history and the risk of coronary heart disease: Comparing predictive

models." European Journal of Epidemiology, vol. 17, no. 7, 2001, pp. 609-620.

OhioLINK Electronic Journal Center, rave.ohiolink.edu/ejournals/article/326423567

Estruch, Ramón et al. “Primary prevention of cardiovascular disease with a Mediterranean diet.”

*The New England journal of medicine* vol. 368,14 (2013): 1279-90.

doi:10.1056/NEJMoa1200303

“Food Environment Atlas Data Access and Documentation Downloads.” *USDA ERS - Data*

*Access and Documentation Downloads*, 2020,

www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/.

George, Brandon et al. “Survival analysis and regression models.” *Journal of nuclear*

*cardiology: official publication of the American Society of Nuclear Cardiology* vol. 21,4

(2014): 686-94. doi:10.1007/s12350-014-9908-2

Goh, Louise Gek, Timothy Welborn, and Satvinder Dhaliwal. "Independent external validation

of cardiovascular disease mortality in women utilizing Framingham and SCORE risk

models: a mortality follow-up study." BMC Women's Health, vol. 14, no. 1, 2014, p. 118.

OhioLINK Electronic Journal Center, doi:10.1186/1472-6874-14-118.

Guo, Chonghui, and Jingfeng Chen. "Big Data Analytics in Healthcare: Data-Driven Methods

for Typical Treatment Pattern Mining." Journal of Systems Science and Systems

Engineering, vol. 28, no. 6, 2019, pp. 694-714. OhioLINK Electronic Journal Center,

doi:10.1007/S11518-019-5437-5.

Heidemann, Christin et al. “Dietary patterns and risk of mortality from cardiovascular disease,

cancer, and all causes in a prospective cohort of women.” *Circulation* vol. 118,3 (2008):

230-7. doi:10.1161/CIRCULATIONAHA.108.771881

Hemingway, Harry et al. “Big data from electronic health records for early and late translational

cardiovascular research: challenges and potential.” *European heart journal* vol. 39,16

(2018): 1481-1495. doi:10.1093/eurheartj/ehx487

Hilleary, Rebecca S, et al. "Gender disparities in patient education provided during patient visits

with a diagnosis of coronary heart disease." Women's Health, vol. 15, 2019, p.

OhioLINK Electronic Journal Center, doi:10.1177/1745506519845591.

Kim, Kyuwoong et al. “Severity of dental caries and risk of coronary heart disease in

middle-aged men and women: a population-based cohort study of Korean adults,

2002-2013.” *Scientific Reports* vol. 9,1 10491. 19 Jul. 2019,

doi:10.1038/s41598-019-47029-3

Lasheras, C et al. “Mediterranean diet and age with respect to overall survival in

institutionalized, nonsmoking elderly people.” *The American journal of clinical nutrition*

vol. 71,4 (2000): 987-92. doi:10.1093/ajcn/71.4.987

Lechner, Katharina et al. “Lifestyle factors and high-risk atherosclerosis: Pathways and

mechanisms beyond traditional risk factors.” *European journal of preventive cardiology*

vol. 27,4 (2020): 394-406. doi:10.1177/2047487319869400

Liu, Yibin, and Heather A Eicher-Miller. “Food Insecurity and Cardiovascular Disease Risk.”

*Current atherosclerosis reports* vol. 23,6 24. 27 Mar. 2021, doi:10.1007/s11883-021-00923-6

Marelli, Ariane J., et al. "Congenital heart disease in the general population: changing prevalence

and age distribution." *Circulation* 115.2 (2007): 163-172.

Mendy, Vincent L et al. “Food Insecurity and Cardiovascular Disease Risk Factors among

Mississippi Adults.” *International journal of environmental research and public health*

vol. 15,9 2016. 15 Sep. 2018, doi:10.3390/ijerph15092016

Moritz, Steffen, and Thomas Bartz-Beielstein. "imputeTS: time series missing value imputation

in R." *R J.* 9.1 (2017): 207.

Moynihan, Paula. "The interrelationship between diet and oral health." *Proceedings of the*

*Nutrition Society* 64.4 (2005): 571-580.

Okoro, Catherine A., MS, et al. "Tooth Loss, and Heart Disease." American Journal of

Preventive Medicine, vol. 29, no. 5, 2005, pp. 50-56. OhioLINK Electronic Journal

Center, doi:10.1016/J.AMEPRE.2005.07.006.

O'Connor, Christopher M et al. “Predictors of mortality after discharge in patients hospitalized

with heart failure: an analysis from the Organized Program to Initiate Lifesaving

Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF).” *American heart*

*journal* vol. 156,4 (2008): 662-73. doi:10.1016/j.ahj.2008.04.030

“Oral Health Data: Explore by Topic.” *Centers for Disease Control and Prevention*, Centers for

Disease Control and Prevention,

nccd.cdc.gov/oralhealthdata/rdPage.aspx?rdReport=DOH\_DATA.ExploreByTopic&islTo

pic=ADT.

Roth, Gregory A, et al. “Global Burden of Cardiovascular Diseases and Risk Factors,

1990-2019: Update From the GBD 2019 Study.” *Journal of the American College of*

*Cardiology* vol. 76,25 (2020): 2982-3021. doi:10.1016/j.jacc.2020.11.010

Sanchez, Paula et al. “Oral health and cardiovascular care: Perceptions of people with

cardiovascular disease.” *PloS one* vol. 12,7 e0181189. 20 Jul. 2017,

doi:10.1371/journal.pone.0181189

Srour, Bernard et al. “Ultra-processed food intake and risk of cardiovascular disease: prospective

cohort study (NutriNet-Santé).” *BMJ (Clinical research ed.)* vol. 365 l1451. 29 May.

2019, doi:10.1136/bmj.l1451

Stacy Westerman, Nanette K. Wenger; Women and heart disease, the underrecognized burden:

sex differences, biases, and unmet clinical and research challenges. *Clin Sci (Lond)* 1

April 2016; 130 (8): 551–563. doi: https://doi.org/10.1042/CS20150586

Stanley, William C, et al. “Does junk food lead to heart failure? Importance of dietary

macronutrient composition in hypertension.” *Hypertension (Dallas, Tex. : 1979)* vol. 54,6

(2009): 1209-10. doi:10.1161/HYPERTENSIONAHA.109.128660

Temple, Norman J. “Fat, Sugar, Whole Grains and Heart Disease: 50 Years of Confusion.”

*Nutrients* vol. 10,1 39. 4 Jan. 2018, doi:10.3390/nu10010039

Vaughan AS, Schieb L, Casper M. Historic and recent trends in county-level coronary

heart disease death rates by race, gender, and age group, United States, 1979-2017. PLOS

ONE 15(7) (2020): e0235839. <https://doi.org/10.1371/journal.pone.0235839>

Wessler, Benjamin S et al. “Clinical Prediction Models for Cardiovascular Disease: Tufts

Predictive Analytics and Comparative Effectiveness Clinical Prediction Model

Database.” *Circulation. Cardiovascular quality and outcomes* vol. 8,4 (2015): 368-75.

doi:10.1161/CIRCOUTCOMES.115.001693

Westerman, Stacy, and Nanette K Wenger. “Women and heart disease, the underrecognized

burden: sex differences, biases, and unmet clinical and research challenges.” *Clinical*

*science (London, England: 1979)* vol. 130,8 (2016): 551-63. doi:10.1042/CS20150586

**Appendices:**

All codes for data cleaning, data wrangling, models, and some of the following graphs were completed in R. They are in the Rmd file titled “DA 401 draft4”. The interactive US map dashboard and some of the following graphs are completed and hosted on both Tableau Desktop and Tableau Server. Both the Tableau Desktop and Tableau servers contain the same materials, I exported the graphs in two different versions just in case the link to the Tableau server has an error. The workbook is titled “DA 401 final manuscript data visuals. Twbx”. The interactive Tableau dashboards and graphs are hosted on Tableau Server and can be accessed via this link:

<https://public.tableau.com/views/DA401fullmanuscriptvizualization/Mapdashboard?:language=en-US&:display_count=n&:origin=viz_share_link>

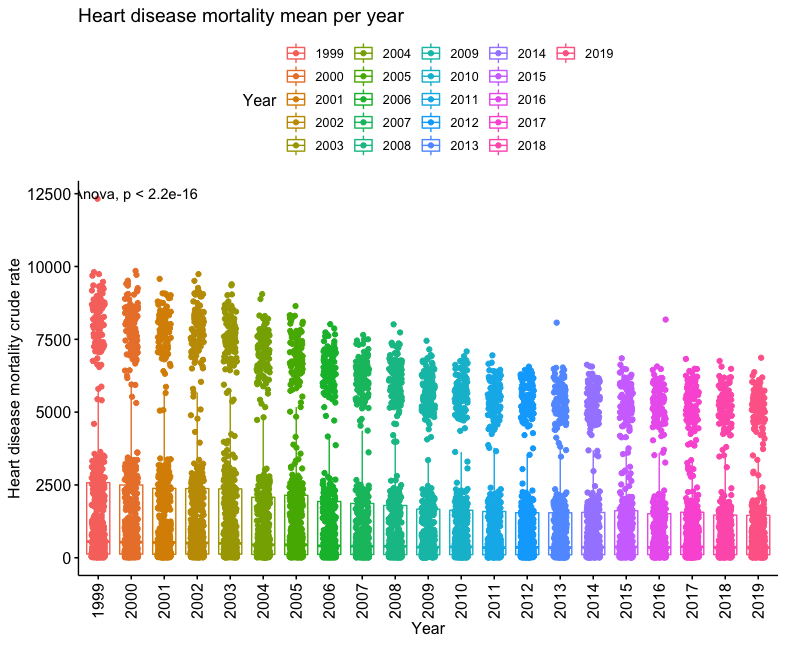


Figure 1: Boxplot of heart disease mortality across years

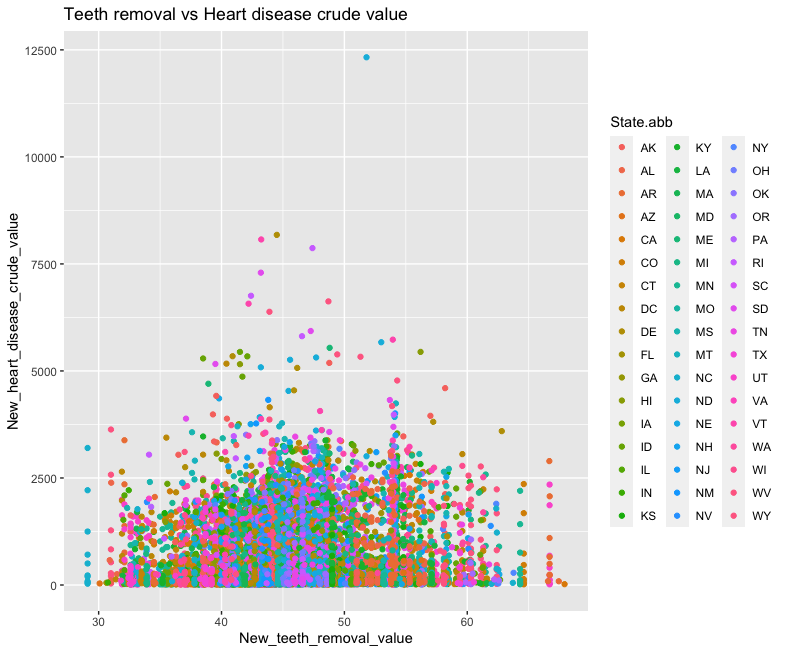


Figure 2: Scatter plot of teeth removal vs. Figure 3: Scatter plot of teeth removal vs.

heart disease death heart disease death with age filter

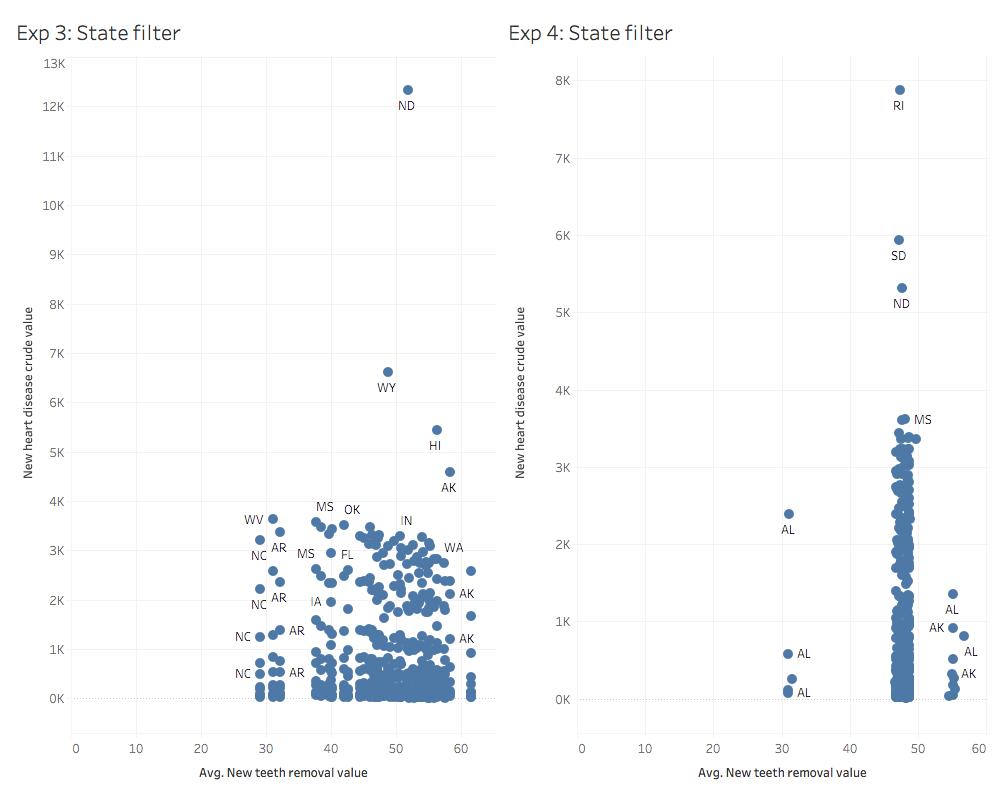


Figure 4: Scatter plots of teeth removal vs. heart disease death by state from 1999 to 2000

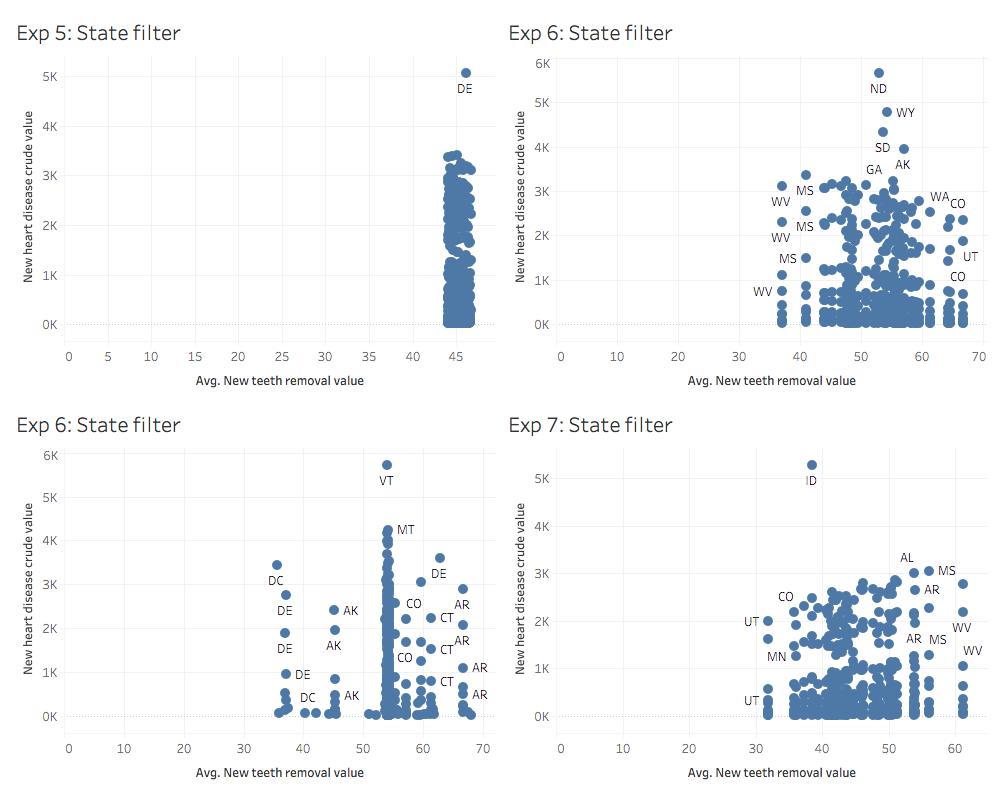


Figure 5: Scatter plots of teeth removal vs. heart disease death by state from 2001 to 2004

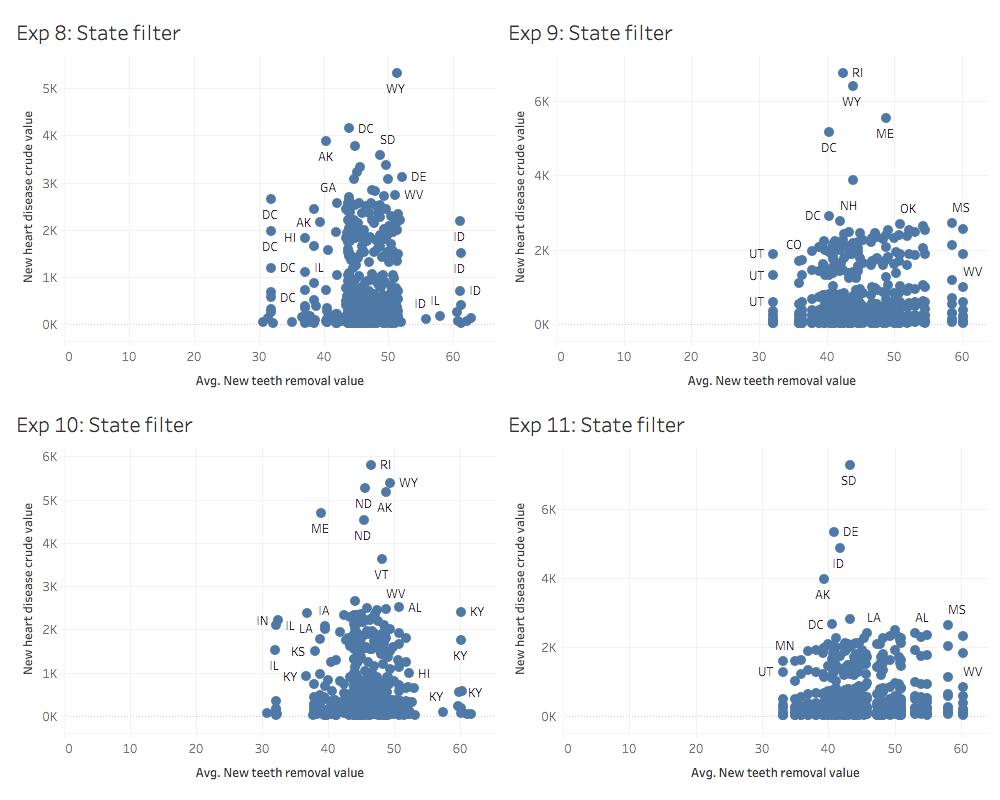


Figure 6: Scatter plots of teeth removal vs. heart disease death by state from 2005 to 208

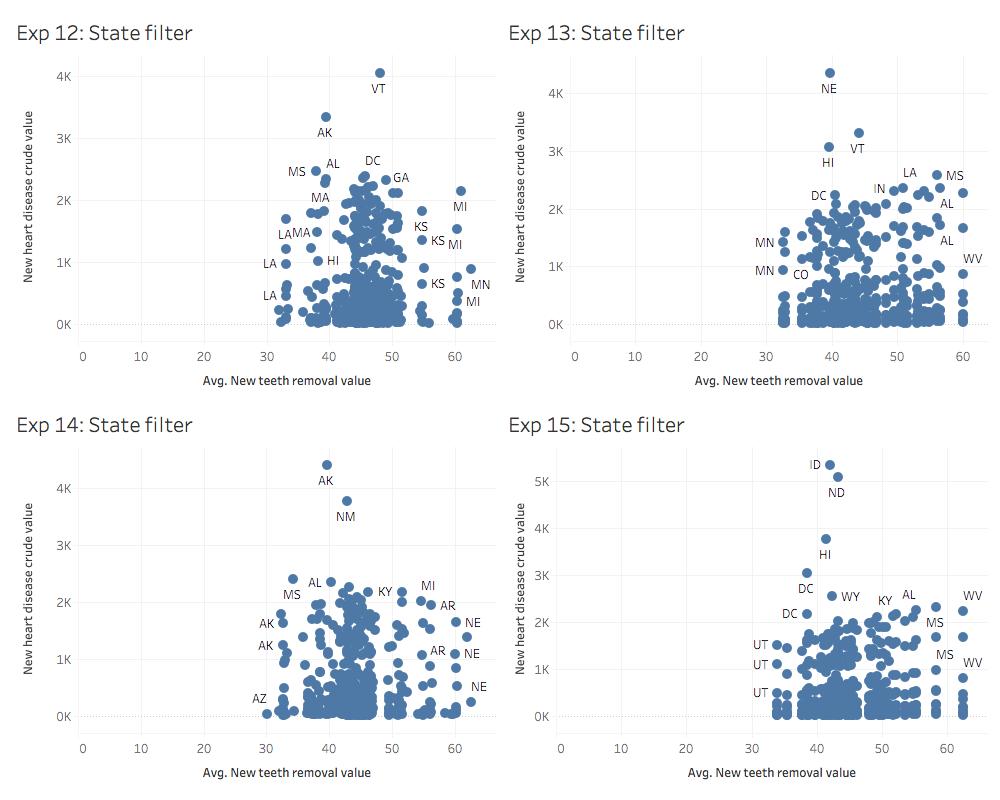


Figure 7: Scatter plots of teeth removal vs. heart disease death by state from 2009 to 2012

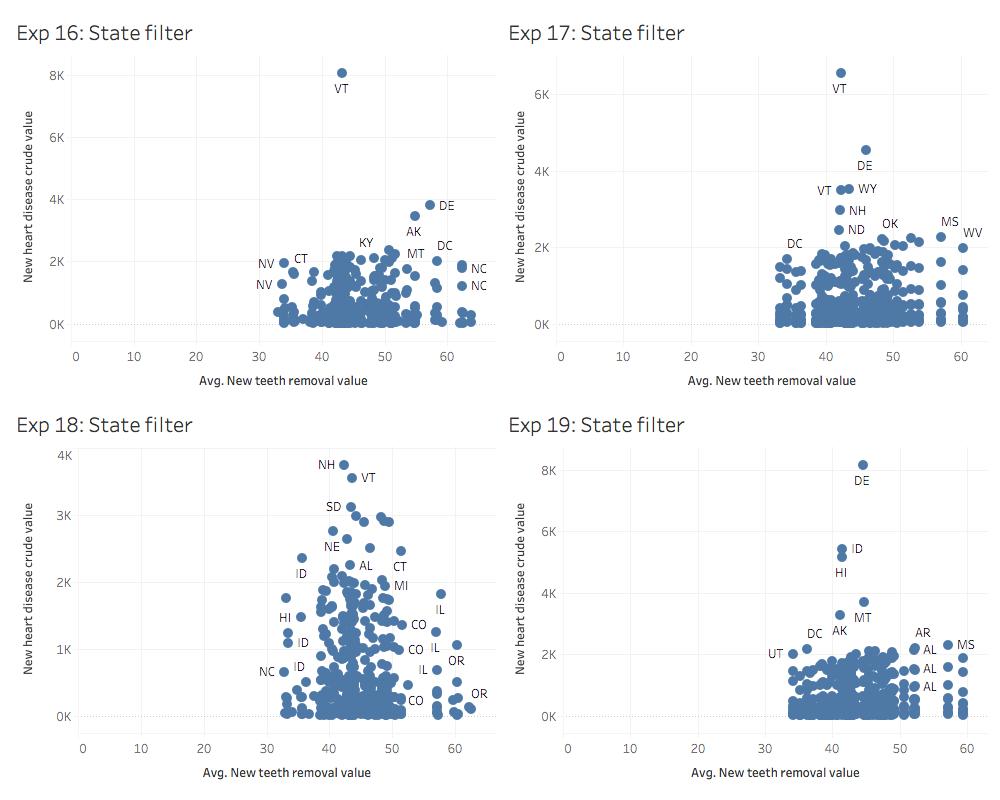


Figure 8: Scatter plots of teeth removal vs. heart disease death by state from 2013 to 2016

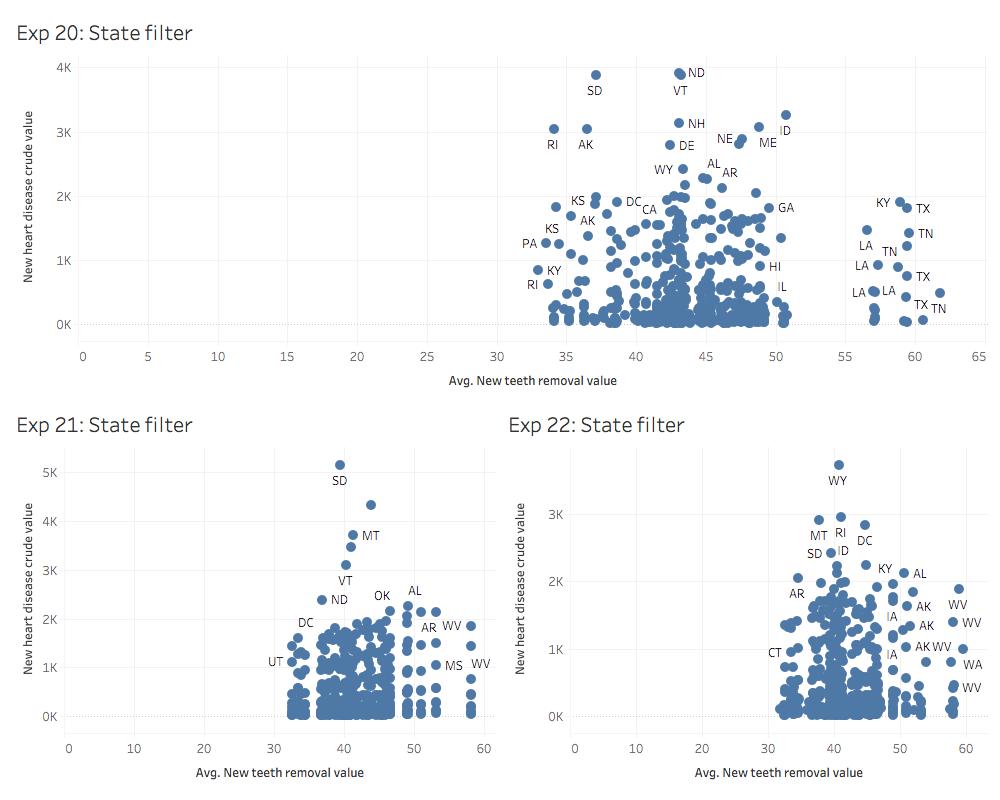
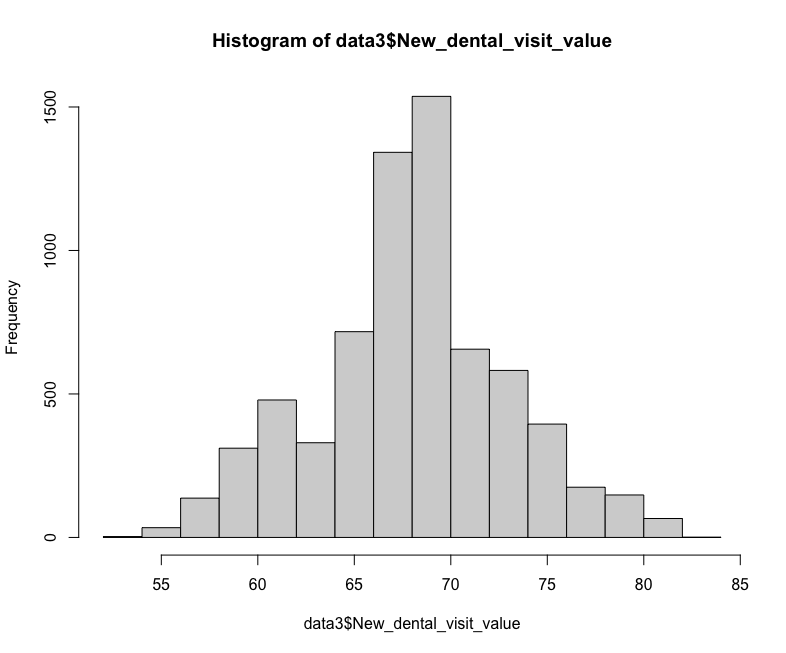
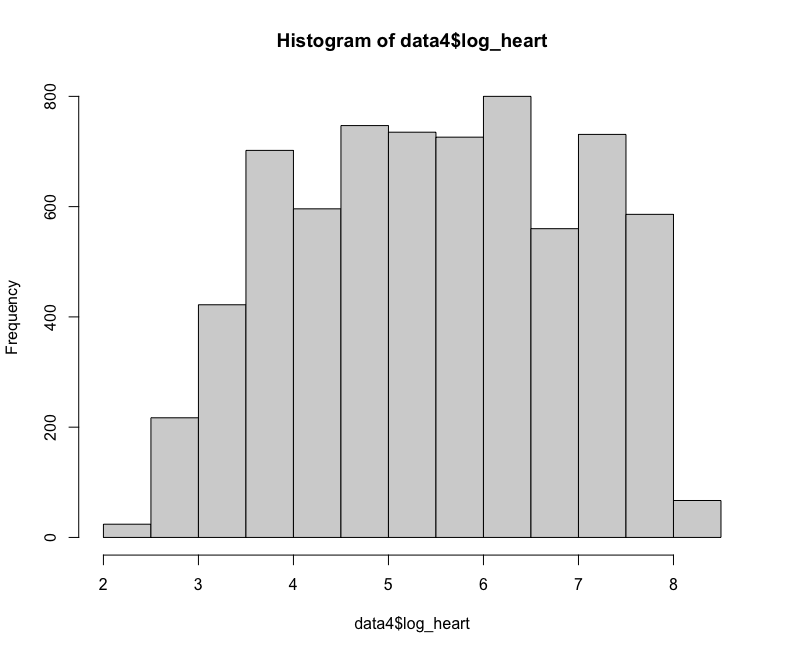
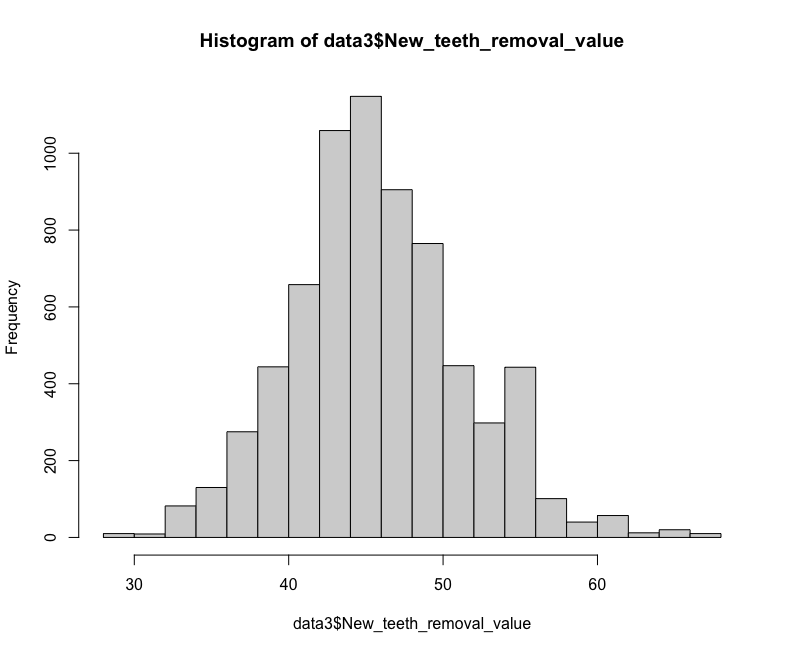
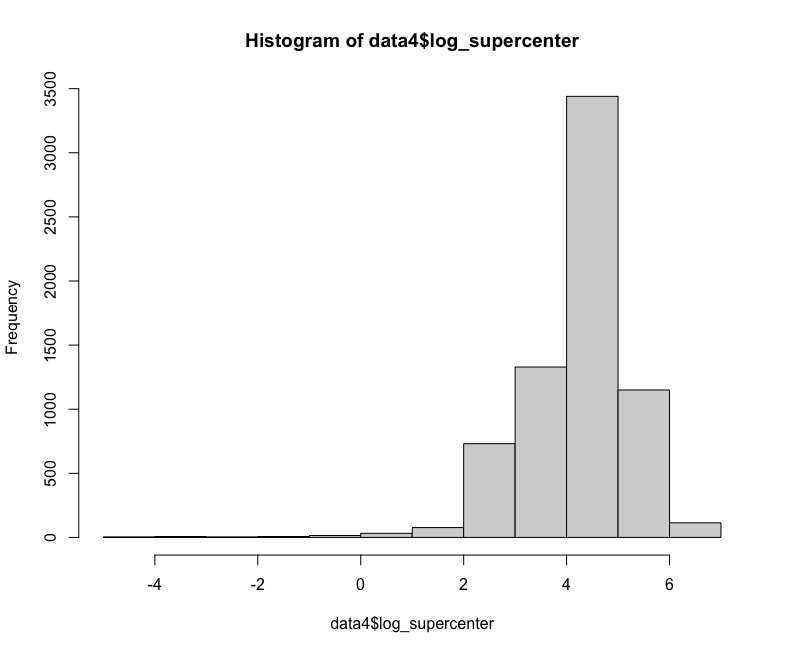
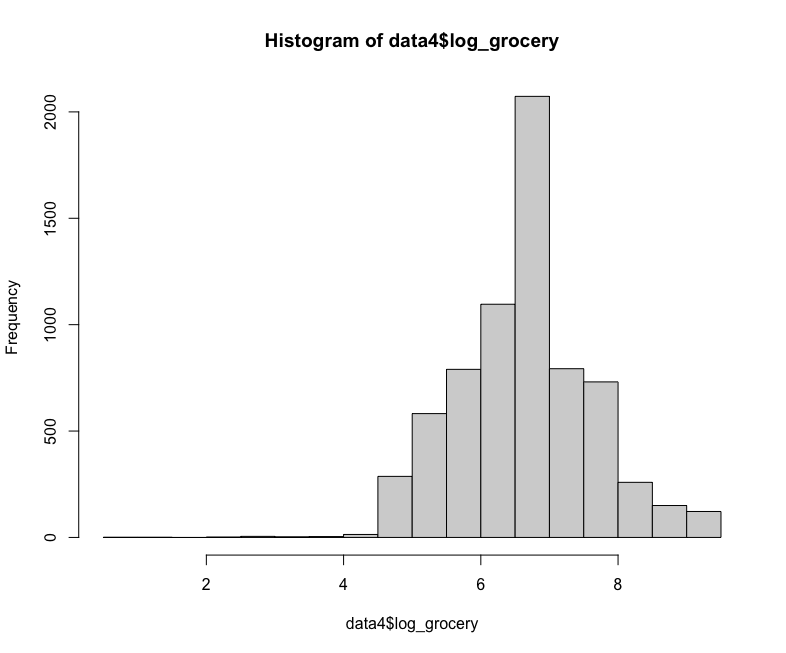
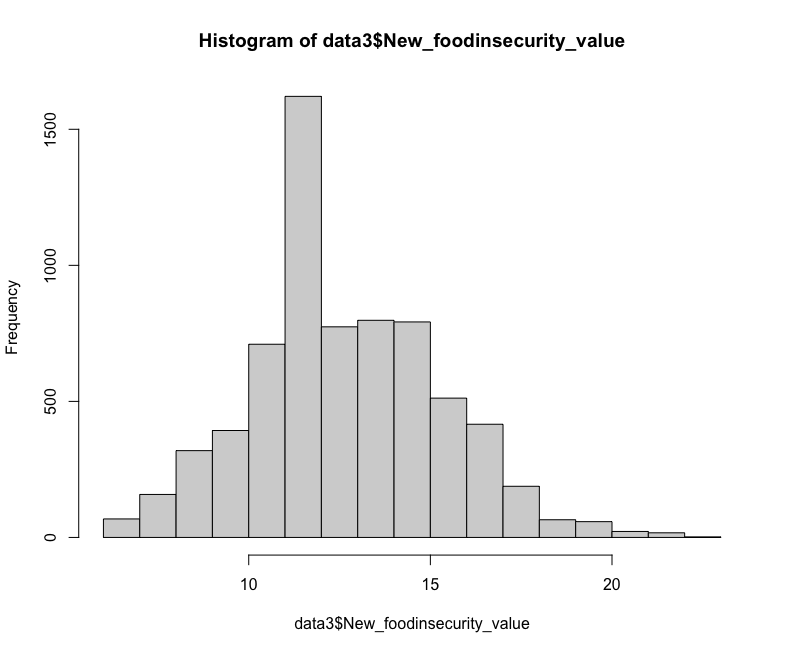
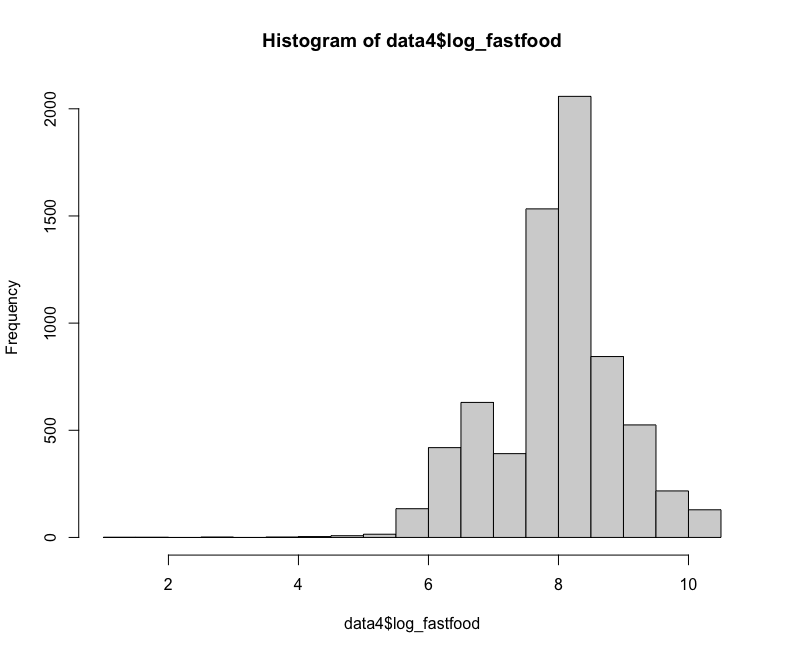
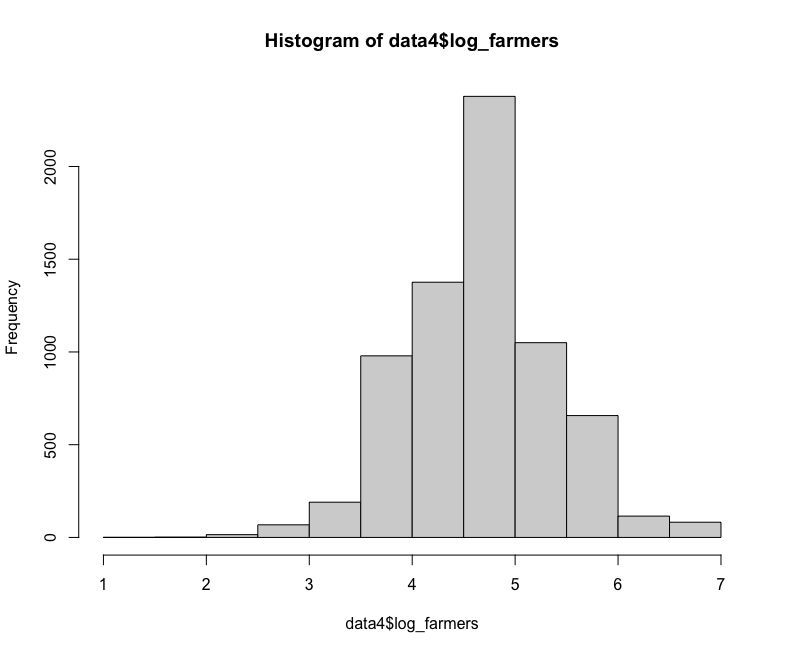
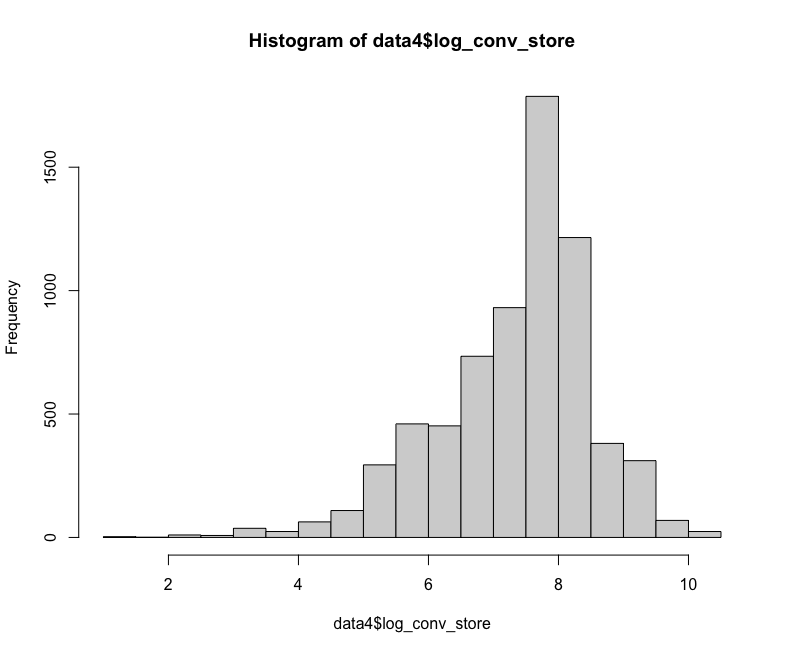


Figure 9: Scatter plots of teeth removal vs. heart disease death by state from 2017 to 2019



Figure 10: Scatter plot of teeth removal vs. heart disease death with state filter





Figures 11-19: Histograms of heart disease mortality and predictors

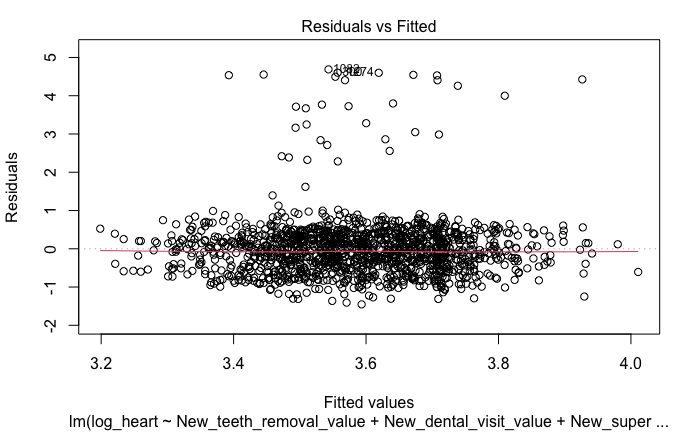
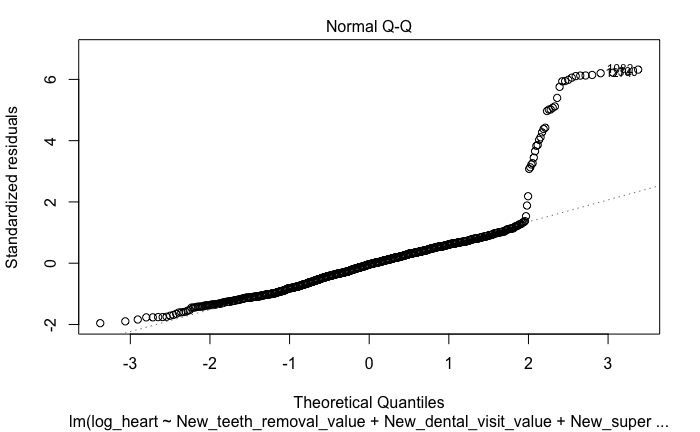
 

Figure 20: Residual and normality plot for fixed effects model of age group 35-44

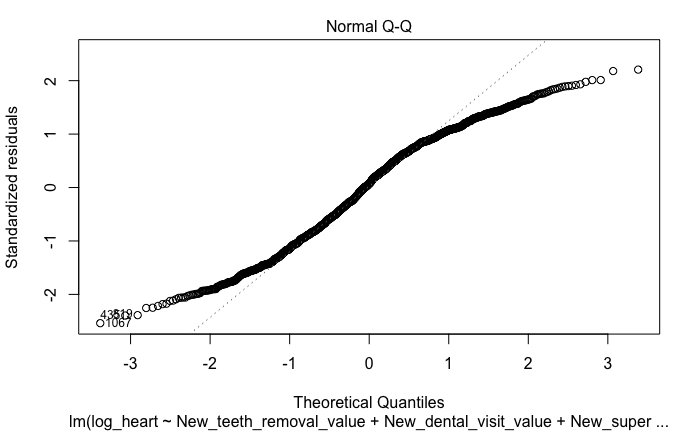
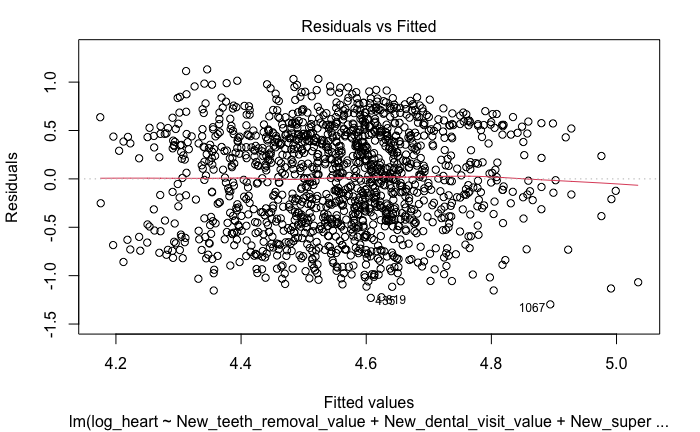


Figure 21: Residual and normality plot for fixed effects model of age group 45-54

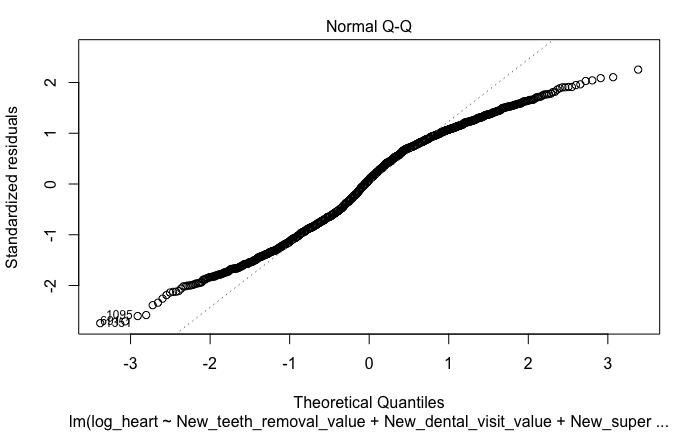
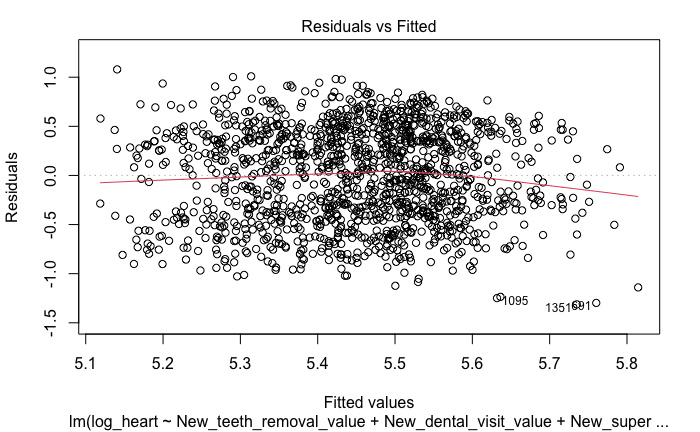


Figure 22: Residual and normality plot for fixed effects model of age group 55-64

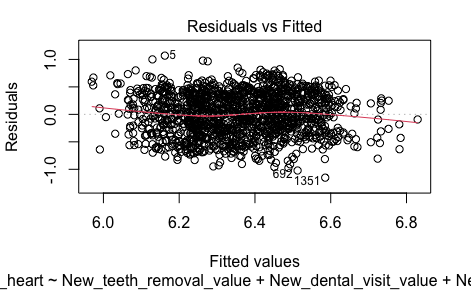
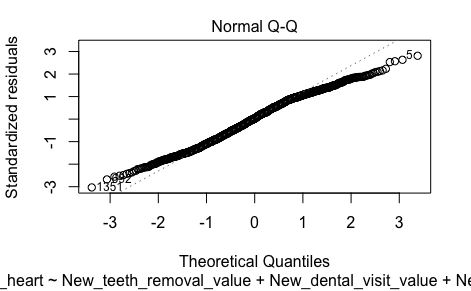
 

Figure 23: Residual and normality plot for fixed effects model of age group 65-74

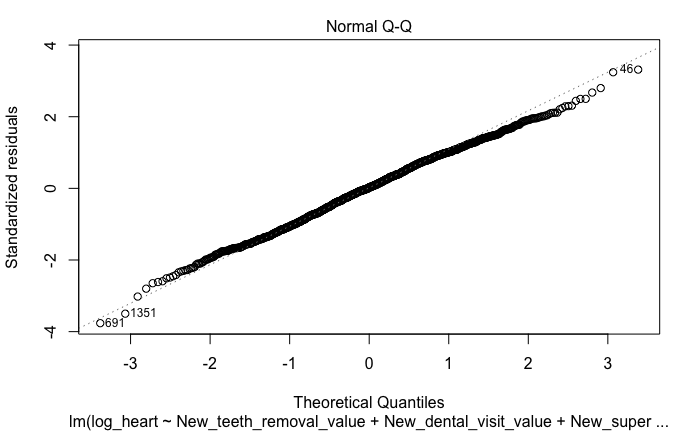
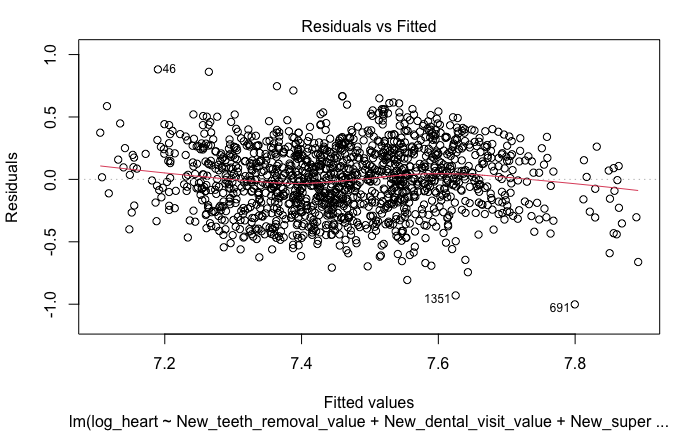


Figure 24: Residual and normality plot for fixed effects model of age group 74-84

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Residual min** | **Residual 1Q** | **Residual median** | **Residual 3Q** | **Residual max** |
| Fixed-effect | -2.72 | -1.02 | -0.00 | 0.99 | 3.02 |

**Table 1:** Shows residual information for the first fixed-effect model after the VIF test