The Influence of Socio-Demographic Variables and Exemption Policies on State-Specific Adult Vaccination Rates

# Executive Summary

Vaccines are a vital component of modern medicine. While many infectious diseases have been contained by the use of vaccines, the seasonal flu is an illness that has not been so easily removed due to the changing nature of the influenza virus. It is recommended by the CDC that individuals receive the flu vaccine annually. With these recommendations in place, there are still millions affected by the flu every year from which thousands die. This statistic comes to show that millions of Americans are going unvaccinated. The intent of this project is to see what variables affect an individual’s decision to get vaccinated against the flu and to determine which variables to target in order to increase rates of influenza vaccination by state. In order to perform this analysis, various variables such as Income, Religion and different races were chosen. After variables were chosen, a multiple linear regression model was generated in order to produce an equation that showed us the relationship between each independent variable and the dependent variable. Interestingly, vaccination exemption policies proved to be significant and showed differing influence depending on region. After this, the clustering method was used in order to cluster variables and states together in order to better categorize findings in order to make optimum recommendations. The clustering method allowed us to form three different clusters of states. In the end, we were able to find that the variables that had the most effect on flu vaccination by state are Religion, Black, Visit, Children, Population55, College, Uninsured and Democratic.

# Introduction

Since the development of the smallpox vaccine in the 1790’s, immunizations have revolutionized healthcare and extended the human lifespan considerably by reducing the prevalence of infectious diseases (Stern et al.). Serving as a preventative measure, vaccines for smallpox and polio have almost eliminated these diseases’ presence in society and other various vaccines have become a foundation for maintaining public health (History of Vaccine). While the vaccine for the measles provides lifelong immunity, influenza virus is not so easily defended against; the CDC recommends annual vaccination for those older than 6 months(About Measles). Every year, from October to March, flu season takes hold and virus activity greatly heightens (Flu Season). In the US from 2016-2017, 29 million people had a symptomatic illness, 14 million individuals required medical visits, 500,000 Americans were hospitalized due to the flu and 38,000 people died as a result (Disease Burden). By opting to receive the vaccine, one can prevent or reduce symptoms of influenza and further, avoid personal productivity loss and medical expenses. The risk of flu is lowered 40 to 60% in the population by the flu vaccination when the vaccine is properly coordinated with the common virus strains of the season (Vaccine Effectiveness). Further testifying to influenza immunization’s utility, the flu vaccination in 2016-2017 resulted in 5,283,410 averted illnesses and 5,217 deaths (Estimated Influenza). Despite these beneficial outcomes, the average annual economic burden due to the flu in the United States is estimated at $3.2 billion to just the healthcare system and further, at $11.2 billion with the inclusion of productivity loss (Putri et al.).

Correspondingly, the extreme example of covid-19 illustrates the devastating economic effects of a lack

of a vaccination in the face of pandemic. The considerable economic impact of influenza and concern for public health has invited governmental intervention but also, has been met with a rising anti-vaccination sentiment.

With origins dating back to the 18th century involving religious protest, resistance against vaccinations has always been present (Benecke et al.). Modern anti-vaxx attitude has emanated from the

debunked publications of Andrew Wakefield who claimed a casual relationship between the MRR vaccine and autism in children in 1998 (Benecke et al.). Consequently, celebrities such as Jenny McCarthy and Robert De Niro have publicly renounced vaccinations, promoting the idea that vaccinations and the development of autism are linked(Benecke et al.). Vaccine hesitancy can be further attributed to societal attitude changes such as reduced trust in medical corporations and their partnerships with the government, and increased interest in natural products and alternative medicines (Siddiqui et al.). Vaccine confidence is dependent upon health care provider endorsement but due to widespread availability of medical information online, which could be misleading, the relationship between a patient and provider has transformed into one with joint decisions and reduced reliance on physician’s expertise. In combination with the abundance of misinformation on social media, these phenomena have had significant results. In 2019, the United States had the highest cases of measles since 1994 which was declared eradicated in 2000 (Patel et al.). In addition, the 2017-2018 flu season was marked with a 6.2% reduction in adult flu vaccination coverage in comparison to the previous year (Estimate of Influenza).

Concern has been concentrated on the vaccines rather than the illnesses themselves.

The substantial socio-economic repercussions of influenza vaccinations and trending

anti-vaccination attitudes justify investigation into state-wide specific influenza vaccination rates. Understanding the factors that influence influenza vaccination coverage can improve public health intervention to minimize burden due to the flu. Further supporting the relevance of our inquiry, the current covid-19 pandemic has highlighted how the absence of vaccination coverage can have profound effects to both the economy and social normalcy.

# Data

Our objective is to identify the key factors that affect the proportion of U.S. citizens’ willingness to get influenza vaccinations. Thus, we will analyze the rate of receiving state-specific flu vaccinations based on these variables using second hand data: income, religion, race, death rate, state expenditure on public health initiatives, hospitals per state, adults with children, medicaid coverage, lack of healthcare visitation due to monetary reasons, political affiliation, education, cellular data plans, state vaccine exemption policy and geographical region.

Our dependent variable flu is the percentage of adults per state who have received a flu vaccine in the past year (2017) and is sourced from the Kaiser Family Foundation based on the CDC’s BRFSS

2013-2018 survey results. Because children are guided by parental decisions, data for adult vaccinations is utilized.

Considering that the ability to pay for a vaccine may affect vaccination decisions, variables regarding economic status are incorporated. Income is the median annual income per household by state in 2017 and is sourced from the Census Bureau’s American Community Survey . In order to represent the population that cannot afford medical costs specifically, the independent variable visit is used to measure the percentage of adults in 2017 who report not seeing a doctor in the last 12 months by state because of cost . This sourced from Kaiser Family Foundation based on the CDC’s BRFSS survey. Because of multicollinearity issues in which independent variables are highly correlated with each other, variables such as poverty rate and unemployment rate are discluded. Further, because insurance can help cover the cost of vaccinations, the variable uninsured, which is the percentage of the state population without insurance in 2017, and the variable medicaid, which is the percentage of the 2017 state population with medicaid insurance, is utilized. Medicaid is a state administered public health insurance program for

low-income eligible adults that apply and can help glean insight on whether having insurance coverage

for low income individuals changes vaccination decisions. Both variables’ data is from the KFF based on the Census Bureau’s ACS.

To consider the influence of access to vaccination, the variable hospitals is included to measure the amount of community hospitals per state in 2017, which is credited to the Kaiser Family Foundation based on the AHA annual survey.

Next, variables are incorporated to recognize the social characteristics that may determine influenza vaccination. Hispanic, Black, and Asian each represent the percentage of the 2017 state population that is of each race. This data is from Kaiser Family Foundation based on the AHA annual survey. Further, vaccinations decisions may be affected by convenience. If a child is receiving a influenza vaccination perhaps due to increased flu spread in school, an adult may opt to receive one as well. Thus, the variable children represents the percent of the adult population by state in 2017 that have children and is sourced from the Kaiser Family Foundation based on the Census Bureau’s ACS. Because older individuals may be prone to more severe consequences of the flu and therefore more likely to receive an influenza vaccination, the variable population 55 and population 65 plus is used to measure the percent of the 2017 state population from 55-64 years old and 65+ years respectively. This data is from the Kaiser Family Foundation based on the Census Bureau’s ACS. Considering that vaccination perception may be affected by education, the variable college is used to represent the percent of the adult population that attained a bachelor’s degree as of 2017. As higher education may be more influential, college was used instead of the percentage of high school graduates. This data originates directly from the Census Bureau based on the ACS.

To acknowledge how perception influences vaccination, variables regarding belief systems are

embodied in this model. Recognizing that certain religions are against vaccinations, religion is a variable that measures the percent of adults by state that report to be highly religious and is sourced from Pew Research Center based on the 2014 Religious Landscape Study. It is justified to use 2014 data for an

approximation as this is the most recent information regarding religiousness by state. Next we consider political affiliation, of which the variable democratic represents the percentage of the 2017 state population that is a democrat or leans democrat. Since America is a two-party country, Republican is excluded because of multicollinearity. This is sourced from the 2017 State Party Affiliation Gallup Poll. In addition, because the perception of danger due to the flu can change a vaccination decision, the variable death is incorporated and is the percentage of total deaths by state that is caused by the flu or pneumonia in 2017. Influenza can cause pneumonia so these two variables are grouped. The number of deaths due to influenza/pneumonia is sourced from the CDC while the total deaths per state 2017 can be cited to Statista. Finally, in consideration that perception can be influenced due to access to information, the variable cellular is utilized and represents the state-specific percentage of the 2017 population that has a cellular data plan. This is directly from the Census Bureau based on the ACS. Variables such as smartphone ownership and internet access were omitted due to high multicollinearity.

Next, public is the state’s budget for health initiatives, which is the dollars committed to public

health and federal dollars relegated to states per person in 2017, and is sourced from American’s Health Rankings of the United Health Foundation. This variable helps us understand if the efforts of public health authorities have encouraged vaccination coverage. To further consider legislation, the dummy variables of mrexemption and philoexemption represent whether or not the state has granted Medical Religious exemptions for school vaccinations or Philosophical exemptions. While influenza vaccine is not part of mandated vaccinations, the exemption policies not only help us understand the strength of state policy but also, the general attitude towards immunizations. Further enriching this understanding, depending on whether or not philosophic exemptions are allowed, monetary means may have a different effect on immunizations. For example, if the belief that immunizations should not be mandatory is more widely accepted, then those with more or less income may have differing choices regarding vaccination

versus states that view immunizations as necessary. Consequently, the interaction between philoexemption and income is incorporated into this model.

Lastly, we control for the South, Midwest and West geographical regions with dummy variables. The Northeast will be automatically excluded from the model. The interactions between philoexemption and south and midwest are included to show if vaccination also varies depending on region and exemption status.

Figure. 1 Descriptive Statistics

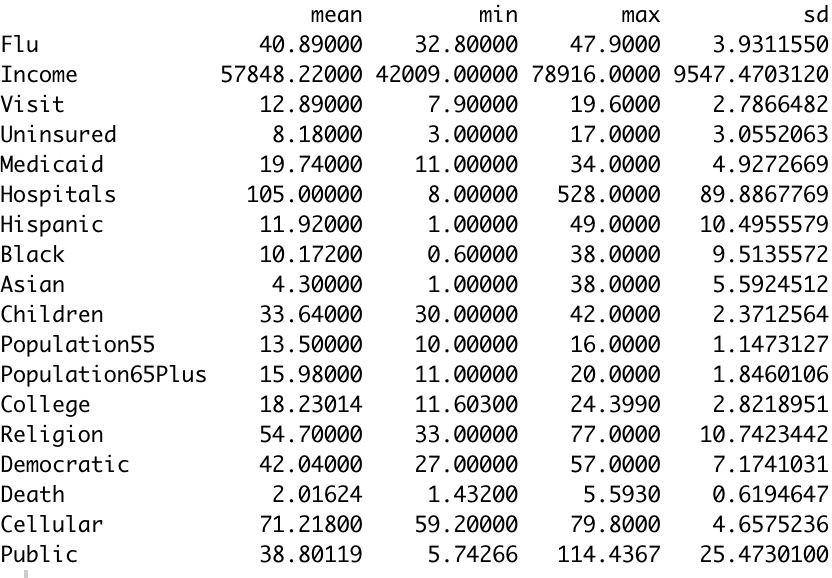


Figure 1 above contains measures of the numeric variables involved in the model to provide an understanding of their values. The dependent variable in question, flu vaccination coverage rate, ranges from 32.80% of Nevada to 47.90% of South Dakota with a mean of 40.89%.

Figure 2. Correlation Visualization

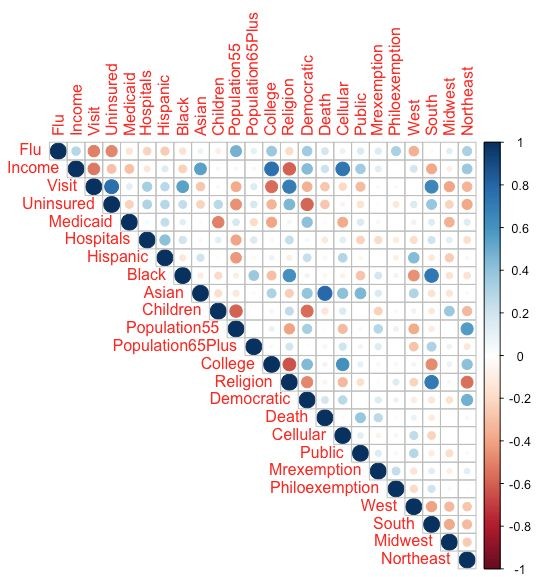
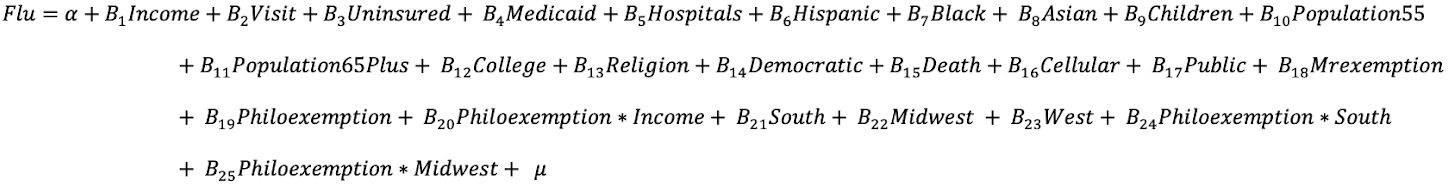


Figure 2 above is a visualization of the correlation between all variables, excluding interactions. As the circles get larger, correlation coefficient is higher with the color indicating positive or negative. Flu is most highly negatively correlated with visit and uninsured, and positively correlated with population 55. While this does provide insight, correlation does not imply causation.

# BI Model

In this report, we will be utilizing a multiple linear regression model to estimate the relationships between the dependent variable, influenza vaccination rate per state, and various socio-economic variables to find out which have the strongest impact.

The regression model to be estimated is as follows:



A regression model was incorporated for our analysis because regression models are well suited to analyze and determine a relationship between independent and dependent models. By using a regression model, we are able to come up with an equation that easily displays the relationship between our multiple independent variables and our dependent variable. The coefficients that are given in the equation produced are the strength between each independent variable and the dependent variable. The coefficient also tells you how the dependent variable is expected to change when the independent variable increases by 1 (while holding all other independent variables constant). Positive coefficients show us that as the independent variable increases, the average of the dependent variable also increases, while negative coefficients tell us that as the independent variable increases, the dependent variable average decreases.

This is useful in our analysis to see what factors (such as race, religion, etc.) could increase or decrease an individual’s decision to vaccinate themselves.

From the different regression models, we decided to utilize linear regression. In comparison to linear regression, logistic regression is mainly used to solve classification problems. A reason that linear regression was chosen over logistic regression is because linear regression allows us to produce results that are continuous (price, age, etc), while logistic regression would give us a categorical, binary output.

In our analysis, we are trying to determine which factors affect the decision to get an influenza vaccine which is why generating a binary response would not be useful in this case.

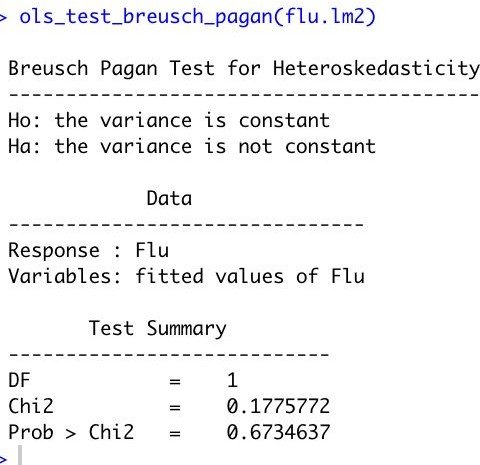
After analyzing the results produced from our linear regression model, we used the clustering method to group together states in order to better categorize our findings and give recommendations. Cluster analysis is vital in identifying homogeneous groups by similar characteristics. In this case, cluster analysis is used to later divide the country into different clusters of states. The clusters place the states into groups with other similar states. This could be useful when implementing future policies, so that policies can be implemented by category instead of by state.

# Alternative Specifications

In order to confirm that all assumptions are met for this model, it was tested for homoscedasticity, normal residuals, multicollinearity and general fit. Because this model is not of a panel data or time-series composition, there are no concerns about serial correlation.

Below is the Breusch-Pagan test for homoscedasticity. With a p-value of 0.6735, we cannot reject the null hypothesis of homoscedasticity so, homoscedasticity is assumed.

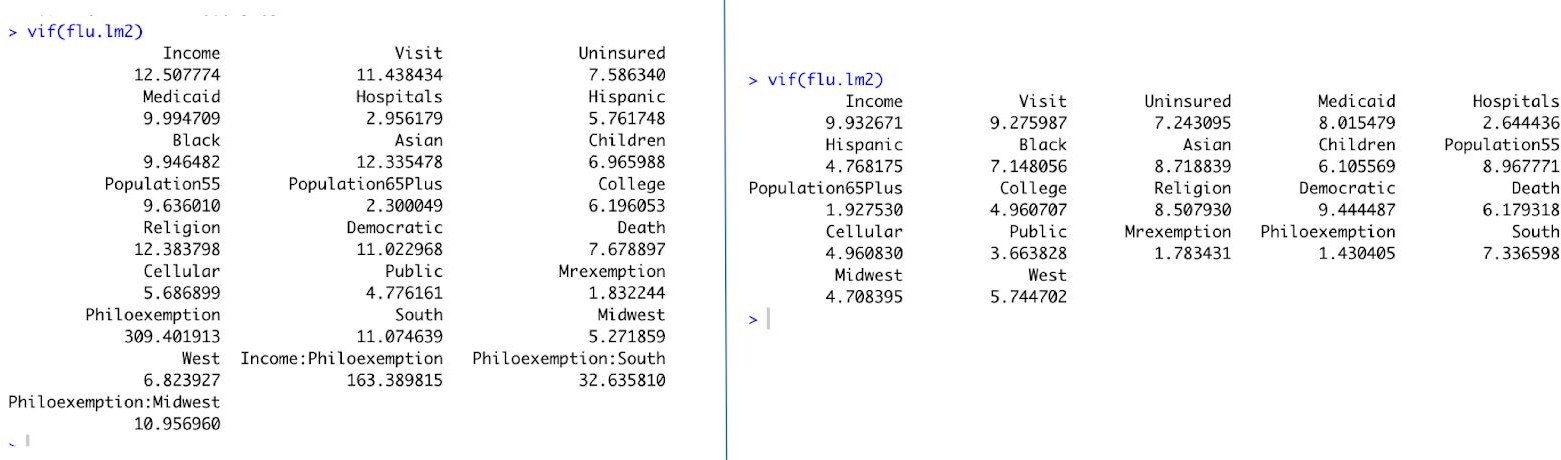
Figure 3. Breusch Pagan Test



To test for multicollinearity, variance inflation factor is utilized. In the left section, vif results are shown for the empirical model. Naturally following, variables involved in the interaction variables show

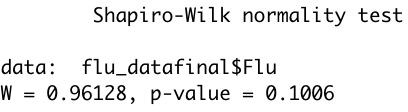
very high multicollinearity. In the right section, vif results are shown for regression without the interaction variables. All values are less than 10.

Figure 4. Variance Inflation Factors with and without Interaction Variables



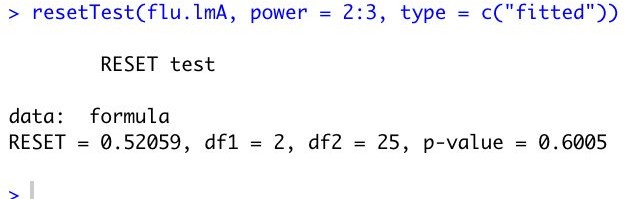
Testing for normality of the independent variable, the shapiro test is utilized. Considering the p-value is above 0.1, we can consider the distribution of data to not be significant from normal distribution.

Figure 5. Normality Test



Finally, the Ramsey RESET is used to test if the model is misspecified and requires non-linear combinations. As shown below, the p-value is insignificant at 0.6005 when testing for quadratic and cubic functions is 0.6005. So according to this diagnostic, the model is adequate.

Figure 6. RESET Test



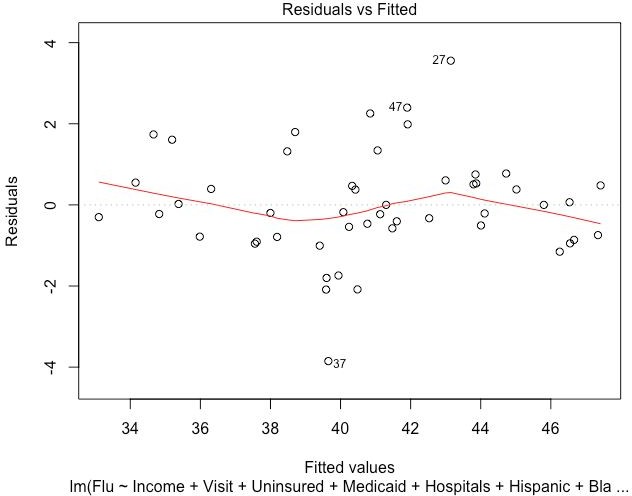
Below in Figure 7, the normal probability plot of residuals graphs fitted residuals of our model against percentiles. As the data is approximately linear, it can be assumed that error terms are approximately normal. The variation in the plot from the line can be partly attributed to the fact that interaction variables with numerical data introduces a non-linear component.

Figure 7. Normal Probability Plot

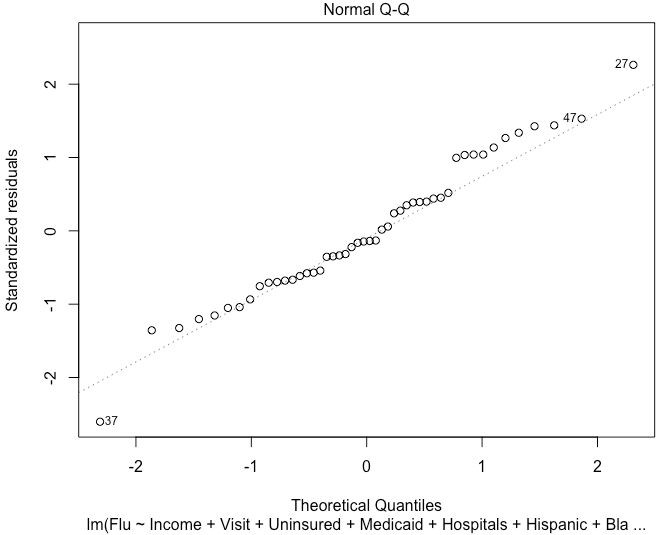


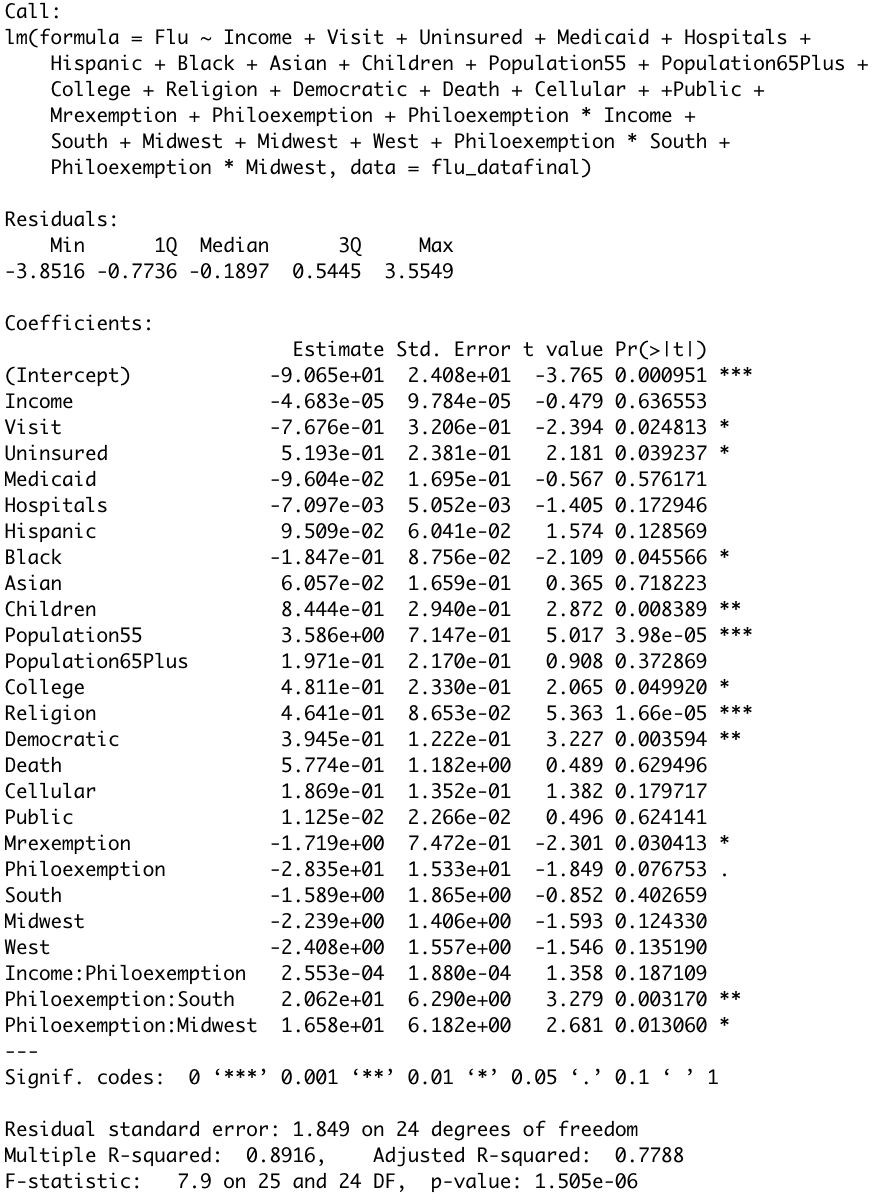
Figure 8 below is a graph of the residuals vs fitted values of our model. The non-linear relationships are due to the interaction variables introduced in the model among income, philosophical exemption, south and midwest. Hence we can infer that variations are not linear between variables and the line is fitted to cater for least square residuals with non-linear effect.

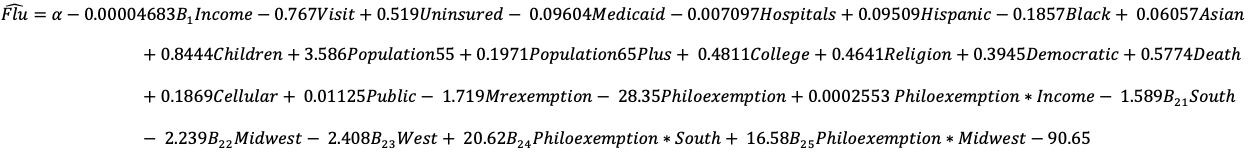
Figure 8. Residuals vs Fitted Plot

# Results

Figure 9 below shows regression results with a corresponding sample regression function.

Figure 9. Regression Results





As shown above, adjusted R squared is 0.7788, meaning that this model explains 77.88% of the variation of data around the mean. This is a relatively high amount. In addition, we have several statistically significant results.

At the 0.001 level of significance, population55 has a coefficient of 3.586 and a p-value of 0.0000398. This means that for a 1% increase in the population from 55-64 per state, holding all other variables constant, the influenza vaccination coverage rate would increase by 3.586%. Secondly, the variable religion has a coefficient of 0.4641 and a p-value of 0.0000166. This indicates that, when holding all other variables constant, for a 1% increase in the state population that identifies as highly religious, the influenza vaccination coverage rate would increase by 0.4641%.

In addition, Children, Democratic, and the interaction of Philoexemption and South are significant at the 0.01 level. For a 1% increase in the percent of adults with children, the influenza vaccination coverage rate increases by 0.8444% when holding all other variables constant. When the percent of the state population that identifies democratic increases by 1%, holding all other variables constant, the influenza vaccination coverage rate increases by 0.3945%. The interaction of philoexemption and south means that, holding all other variables constant, the effect of being a southern state with philosophical exemptions increases the vaccination rate by 90.65% compared to the effect of philosophical exemptions in other geographical regions. This means that philosophical exemption has a differing influence on influenza vaccination with relation to geographic region.

Next, the variables Visit, Uninsured, Black, College, Mrexemption and the interaction of Philoexemp and Midwest are significant at the 0.05 level. For a 1% increase in adults reporting having not visited the hospital due to money, the influenza vaccination coverage rate decreases by 0.7676% when holding all other variables constant. For a 1% increase in the state population that is uninsured, the influenza vaccination coverage rate increases by 0.5193% when holding all other variables constant. For a 1% increase in the state population that is black, the influenza vaccination coverage rate decreases by

0.1847% when holding all other variables constant. For a 1% increase in the state population that has a bachelor’s degree, the influenza vaccination coverage rate increases by 0.4811% when holding all other variables constant. The coefficient for the dummy variable of Mrexemption means that states that allow medical or religious exemptions, in comparison to states without exemptions, have a 1.719% decrease in flu vaccination.The interaction of philoexemption and Midwest means that, for holding all other variables constant, the effect of being a midwest state with philosophical exemptions increases the vaccination rate by 16.58% compared to effect of having philosophical exemptions in other regions. This means that philosophical exemption has a differing influence on influenza vaccination with relation to geographic region.

Lastly, the dummy variable of philoexemption is significant at the 0.1 level. This means that states that allow for philosophical exemption for vaccines, in comparison to states without exemptions, have 28.35% decrease in influenza vaccination coverage, holding all other variables constant.

# Conclusions

Our results reaffirm the importance of many socio-economic variables. As visit was significant and lowered vaccination rates, it is evident that the specific financial ability to go to a healthcare facility is important. By contrast, the variable uninsured provided unusual results which indicate that as the uninsured rate increases, vaccinations increase. For the population that are uninsured, hospital costs due to influenza illness would be much greater so they may be more avoidant and in turn, receive the influenza vaccine. While race does affect vaccination rates, the age demographic is particularly interesting: the population from 55-64 is statistically significant in contrast to the population 65 plus who would be more susceptible to serious consequences of the flu. With this in mind, it could be true that the age range

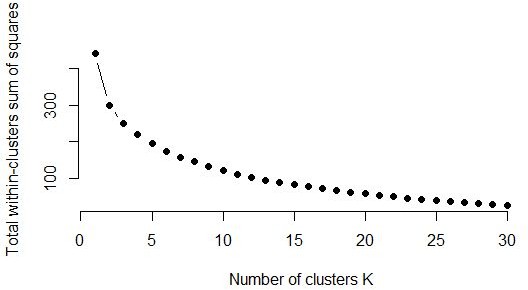
55-64 contains individuals that are more mobile and therefore, more able to access healthcare to receive vaccinations in comparison to the population 65 plus. Further, the percentage of adults with children

proved to be key. Perhaps due to increased social interaction of familial obligations and mandatory schooling, adults with children receive more vaccinations not only because they perceive flu risk as higher, but also opt in at the opportunity with their family. Next, belief systems do have important consequences to vaccination rates. Religiosity positively affected influenza vaccination rates which was unpredicted. Similarly, democratic predicts a positive increase in seasonal influenza vaccination. Perhaps these belief systems indicate schools of thought that are involved in the decision-making of receiving an influenza vaccine. Not only does the belief of vaccine effectiveness change opinion but also, the perception of risk of getting influenza, the worry due to risk, and consideration of side effects influence decision making(Santibanez et al). Thus, these variables and others may have a direct affiliation with these deliberations. For instance, in a 2012 study by the CDC about socio-demographic differences in opinions, it was found that less blacks believed they were susceptible to influenza than other racial groups. Our results demonstrate that the increase of black representation does indeed decrease influenza vaccination. Finally, state mandated vaccine exemption policies have a considerable impact on vaccinations. While influenza is an optional and seasonal vaccination and is not controlled by public policy, these categorical variables indicate a state-directed sentiment towards immunizations. Considering that medical and religious exemption policies mainly affect children due to public school prevalence, the result of decreasing vaccination rates in adults is a critical finding. These policies targeted towards children that create mandatory immunizations have a considerable consequence towards the influenza vaccinations of adults which is optional. Public pressure for vaccination rates could be minimized in the areas and therefore, impact decision making. Furthermore, the indication that philosophical exemptions decrease vaccination rates, but increases vaccination rate in different geographic regions, emphasizes the effect of attitude towards immunizations in determining vaccination coverage.

# Recommendations

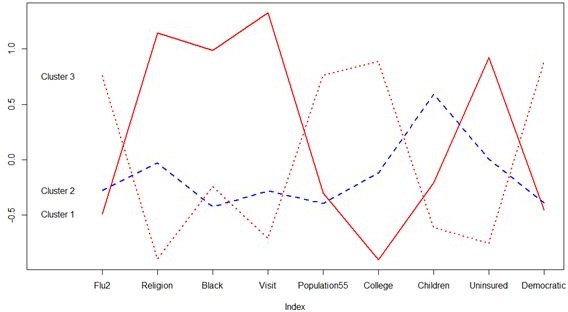
Our model is limited as the coefficient of the regressor would only give us the impact of the variables on all the states. With K-means clustering, we can examine further to help us give recommendations for each state considering important factors.. Forming clusters with similarity can give a generalized idea on a group of states. We consider only significant numeric variables for clustering namely Flu, Religion, Black, Visit, Children, Population55, College, Uninsured and Democratic.

Figure 10. Elbow Graph



As shown above in figure 10, the elbow method is utilized to determine the optimum number of clusters. Maximum cluster option is set to 30. The graph is a plot with the number of clusters on the x-axis and the total within the cluster sum of squares on the y-axis. The point on the elbow of the arm approximates to value 3 which is optimum. Point 3 corresponds to a low sum of squared errors without overfitting the data.

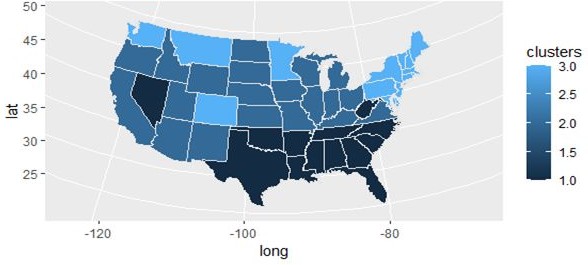
Figure 11. Cluster Characteristics



Data is normalized before running the K-Mean clustering algorithm. Running the algorithm gives three clusters and the fluctuation of each variable across the cluster is plotted.

Further, the clusters are mapped onto the US to show cluster classification.

Figure 12. US Map



Cluster Analysis is as follows:

Cluster 1 : States belonging to this cluster have low flu vaccination rate and very low college graduation rate. High uninsured rate, college graduates, and percent not visiting the hospital due to money needs to be addressed to improve the Flu Vaccination among its citizens as these are indicators of low vaccination rate. This group has a smaller number of people above 55 which adds a positive effect on the Vaccination. These states should focus on improving their economic environment as they seem to be marked by low-income characteristics.

Cluster 2 : This group contains contrasting effects with a high percentage of population with children, less aged people and black people . Democrats are also less in number in these states. Most characteristics are intermediate. This results in a vaccination rate that is mediocre. Perhaps implementing policy reforms for healthcare access and vaccine awareness campaigns to change perception can improve vaccination coverage.

Cluster 3 : On a positive note, this cluster has the maximum vaccination coverage rate so these characteristics must be positively impacting the decisions for vaccination. While most characteristics affect vaccination coverage rate in a positive way such as low uninsured, this cluster is marked by low religiosity and children which has a negative impact. These states must maintain their economic climate and demographic to preserve their impressive vaccination standing.

# R Code

##original spreadsheet with 49 variables #library(readxl)

#project\_data\_r3 <- read\_excel("R/project\_data\_r3.xlsx") #View(project\_data\_r3)

flu\_data <- project\_data\_r3

flu\_data <- flu\_data[ ,-c(1)] #remove state names

##final spreadsheet with 24 variables library(readxl)

project\_data\_final <- read\_excel("R/project\_data\_final.xlsx") View(project\_data\_final)

flu\_datafinal <- project\_data\_final

flu\_datafinal <- flu\_datafinal[ ,-c(1)] #remove state names

*#For Data Description Table*

#data description data.frame(mean=sapply(flu\_datafinal, mean),

min = sapply(flu\_datafinal, min), max = sapply(flu\_datafinal, max), sd = sapply(flu\_datafinal, sd), class = sapply(flu\_datafinal, class))

*#For Correlation Visualization* # correlation visualization #install.packages("ggplot2") library(ggplot2)

cormatrix<-signif(cor(flu\_data),2) cormatrix

corrplot::corrplot(cormatrix, type = "upper")

*#Install Various Packages* library(tidyverse) library(caret) #install.packages("olsrr")

#install.packages("fRegression") library(fRegression) library(olsrr) #install.packages("car") library("car")

*#Final Model*

flu.lm2 <-lm(Flu~ Income + Visit + Uninsured + Medicaid + Hospitals + Hispanic + Black + Asian + Children + Population55 + Population65Plus + College + Religion + Democratic + Death + Cellular + +

Public + Mrexemption + Philoexemption + Philoexemption\*Income + South + Midwest + Midwest + West + Philoexemption\*South + Philoexemption\*Midwest, data = flu\_datafinal)

summary(flu.lm2)

*#test for homoskedasticity*

ols\_test\_breusch\_pagan(flu.lm2)

*#without interactions for testing multicollinearity*

flu.lm2 <-lm(Flu~ Income + Visit + Uninsured + Medicaid + Hospitals + Hispanic + Black + Asian + Children + Population55 + Population65Plus + College + Religion + Democratic + Death + Cellular + Public + Mrexemption + Philoexemption + South + Midwest + Midwest + West, data = flu\_datafinal)

*#test for multicollinearity*

vif(flu.lm2)

*#Specification test*

resetTest(flu.lmA, power = 2:3, type = c("fitted")) resetTest(flu.lmA, power = 2, type = c("fitted")) resetTest(flu.lmA, power = 3, type = c("fitted"))

*#Normality testing of dependent variable*

qqPlot(flu\_data$Flu2) #all points fall approximately along this reference line-> we can assume normality. shapiro.test(flu\_datafinal$Flu) #P-value above benchmark, assume normality

*#plot residuals vs percentiles*

a<- plot(flu.lm2) a

#*plot fitted vs residuals*

plot(lm(Flu~ Income + Visit + Uninsured + Medicaid + Hospitals + Hispanic + Black + Asian + Children

+ Population55 + Population65Plus + College + Religion + Democratic + Death + Cellular + + Public + Mrexemption + Philoexemption + Philoexemption\*Income + South + Midwest + Midwest + West + Philoexemption\*South + Philoexemption\*Midwest, data = flu\_datafinal))

abline(flu.lm2)

############Cluster##########

clust <- read.csv("cluster.csv") row.names(clust) <- clust[,1] clust.name <- clust[,1]

clust <- clust[,-c(1)]

#########Data Normalization####### clust.norm.df <- sapply(clust, scale)

#######Elbow graph##########

set.seed(10) k.max <- 30

data <- clust.norm.df sum\_square <- sapply(1:k.max,

function(k){kmeans(clust.norm.df, k, nstart=50,iter.max = 15 )$tot.withinss})

sum\_square

plot(1:k.max, sum\_square,

type="b", pch = 19, frame = FALSE, xlab="Number of clusters K",

ylab="Total within-clusters sum of squares")

#########K-Mean##########

set.seed(10)

km4 <- kmeans(clust.norm.df, 3) km4

########## Cluster Plot#############

plot(c(0), xaxt = 'n', ylab = "", type = "l",

ylim = c(min(km4$centers), max(km4$centers)), xlim = c(0, 9)) axis(1, at = c(1:9), labels = names(clust))

for (i in c(1:3))

lines(km4$centers[i,], lty = i, lwd = 2, col = ifelse(i %in% c(1, 3, 5),

"red", "blue"))

text(x = 0.2, y = km4$centers[, 1], labels = paste("Cluster", c(1:3)))

##########Plot on the US map########## install.packages("maps")

library(maps) library(usmap) library(ggplot2) library(dplyr)

us\_states <- map\_data("state")

States <- clust.name clusters <- km4$cluster clusters

pdata.df <- data.frame(States,clusters) pdata.df$region <- tolower(pdata.df$States) us\_mapplot <- left\_join(us\_states, pdata.df) head(us\_mapplot)

p <- ggplot(data = us\_mapplot, aes(x = long, y = lat,

group = group, fill = clusters))

p + geom\_polygon(color = "gray90", size = 0.1) + coord\_map(projection = "albers", lat0 = 39, lat1 = 45)

#linear regression options

#linear regression with ALL variables besides northeast

flu.lm2 <- lm(Flu2~ Income + Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death + Graduation + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + Republican + College + Male + Female + Divorced + Married + NeverMarried + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6505

## Check for multicollinearity, want VIR below 10 as benchmark

vif(flu.lm2) #lots of high numbers

#linear regression with all variables besides northeast, male, female, divorced, nevermarried (just married)

flu.lm2 <- lm(Flu2~ Income + Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death + Graduation + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + Republican + College + Married + West + South + Midwest, data = flu\_data)

summary(flu.lm2) #adjR 0.6188

vif(flu.lm2) #republican extremely high so remove it since we have democratic

#linear regression with allvariables besides northeast, male, female, divorced, nevermarried (just married), republican

flu.lm2 <- lm(Flu2~ Income + Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death + Graduation + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Married + West

+ South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.607

vif(flu.lm2) #college VIR 12 and Income 14 so remove graduation

#linear regression with allvariables besides northeast, male, female, divorced, nevermarried (just married), republican, graduation

flu.lm2 <- lm(Flu2~ Income + Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Married + West + South + Midwest, data = flu\_data)

summary(flu.lm2) #adjR 0.622

vif(flu.lm2) #college VIR now 7.8 and Income 14. remove income?

#linear regression with allvariables besides northeast, male, female, divorced, nevermarried (just married), republican, graduation

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Married + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6274

vif(flu.lm2) #visit and democratic barely over 10 ols\_test\_breusch\_pagan(flu.lm2) # it is homoskedastic

#linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + West + South + Midwest, data = flu\_data)

summary(flu.lm2)

#adjR 0.6355 #increased R2! vif(flu.lm2)

#linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation -COLLEGE INSTEAD

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + Graduation + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6352 #reduced R2! go back to including college and not hs graduation rates

##MODEL A

#linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation

flu.lmA <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + West + South + Midwest, data = flu\_data)

summary(flu.lmA) #adjR 0.6355

#test for proper model fit #test for multicollinearity

vif(flu.lmA) #visit and democratic barely over 10

#test for heteroskedasticty ols\_test\_breusch\_pagan(flu.lmA) #is is homoskedastic

#test for non-linear fit with ramsey reset tst

<http://fmwww.bc.edu/EC-C/F2014/2228/ECON2228_2014_8.slides.pdf> resetTest(flu.lmA, power = 2:3, type = c("fitted"))

resetTest(flu.lmA, power = 2, type = c("fitted")) resetTest(flu.lmA, power = 3, type = c("fitted"))

## add internet/computer options

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD INTERNET

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Internet + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6331 REDuced Rsquared vif(flu.lm2) #internet VIF 12

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD computer

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Computer + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6272 REDuced Rsquared vif(flu.lm2) #computer VIF 13

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD smartphone

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Smartphone + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6353 almost same r squared vif(flu.lm2) #smartphone VIF 12

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD Nocomputer

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + NoComputer + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6223 lowered R squared a little

vif(flu.lm2) #NoComputer VIF 4 but democratic now 11

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD Cellular

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6411 RAISED R2

vif(flu.lm2) #medicaid now 11 but VIF cellular ok

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD NoInternet

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + NoInternet + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.6331 R2 a little lower vif(flu.lm2) #Nointernet12

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD Cellular REMOVE Medicaid

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data)

summary(flu.lm2)

#adjR 0.6307 lowered R2 vif(flu.lm2) #visit barely above 10

#MODEL A:linear regression with all variables besides northeast, male, female, divorced, nevermarried, married, republican, graduation. ADD Cellular REMOVE VISIT

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data) summary(flu.lm2)

#adjR 0.5706 lowered R2 a LOT! t vif(flu.lm2) #visit barely above 10

# so add cellular to model A to make model B

#MODEL B:linear regression with all variables PLUS CELLULAR without northeast, male, female, divorced, nevermarried, married, republican, graduation.

flu.lmB <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data) summary(flu.lmB)

#adjR 0.6411 RAISED R2

vif(flu.lmB) #medicaid 11.62, Visit 10.61 ols\_test\_breusch\_pagan(flu.lmB) #passes p = 0.37

# if we remove medicaid

flu.lm2 <- lm(Flu2~ Poverty + Religion + Uninsured + Hispanic + Black + Asian + Unemployed + Death

+ Public + Hospitals + Children + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data)

summary(flu.lm2)

#adjR 0.6307 lowered R2

vif(flu.lm2) #visit 10.45 not much change

ols\_test\_breusch\_pagan(flu.lm2) # p is GREATLY reduced to 0.14- makes it more heteroskedastic?? weird. Decide if significant

#remove uninsured

flu.lm2 <-lm(Flu2~ Poverty + Religion + Hispanic + Black + Asian + Unemployed + Death + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp

+ Democratic + College + Cellular + West + South + Midwest, data = flu\_data) summary(flu.lm2) #R 0.6377

vif(flu.lm2) #all below 10 ols\_test\_breusch\_pagan(flu.lm2)

#replace with death2

flu.lm2 <- lm(Flu2~ Poverty + Religion + Hispanic + Black + Asian + Unemployed + Death2 + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp

+ Democratic + College + Cellular + West + South + Midwest, data = flu\_data) summary(flu.lm2)#R 0.6528

vif(flu.lm2) #below 10 ols\_test\_breusch\_pagan(flu.lm2)

#removed death bc coefficient so high

flu.lm2 <-flu.lm2 <-lm(Flu2~ Poverty + Religion + Hispanic + Black + Asian + Unemployed + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp

+ Democratic + College + Cellular + West + South + Midwest, data = flu\_data) summary(flu.lm2) #R 0.6283

vif(flu.lm2) #all below 10

#replace with new death

flu.lm2 <-lm(Flu2~ Poverty + Religion + Hispanic + Black + Asian + Unemployed + NewDeath + Public

+ Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data) summary(flu.lm2) #R 0.6275 #religion, black, public, medicaid, visit, pop75, democratic vif(flu.lm2) #all below 10

ols\_test\_breusch\_pagan(flu.lm2) resetTest(flu.lmA, power = 2:3, type = c("fitted"))

#replace with death2 ##MODEL C ##April 18

flu.lm2 <-lm(Flu2~ Poverty + Religion + Hispanic + Black + Asian + Unemployed + Death2 + Public + Hospitals + Children + Medicaid + Visit + Population\_aff\_flu + Population55 + Mrexemp + Philoexemp

+ Democratic + College + Cellular + West + South + Midwest, data = flu\_data)

summary(flu.lm2) #R 0.6528 #religion, black, asian, death2, public, medicaid, visit, pop75, democratic vif(flu.lm2) #all below 10

ols\_test\_breusch\_pagan(flu.lm2) resetTest(flu.lmA, power = 2:3, type = c("fitted"))

#remove pop aff flu

flu.lm2 <-lm(Flu2~ Poverty + Religion + Hispanic + Black + Asian + Unemployed + Death2 + Public + Hospitals + Children + Medicaid + Visit + Population55 + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data)

summary(flu.lm2) #R 0.6651 #religion, black, asian, death2, public, medicaid, visit, pop75, democratic vif(flu.lm2)

ols\_test\_breusch\_pagan(flu.lm2)

# MODEL D: removed unemployment and pop\_aff\_flu, replaced poverty with income and added interaction variable bw income and philoexemp

flu.lm2 <-lm(Flu2~ + Philoexemp\*Income + Uninsured + Income + Religion + Hispanic + Black + Asian

+ Death2 + Public + Hospitals + Children + Medicaid + Visit + Population55 + Mrexemp + Philoexemp

+ Democratic + College + Cellular + West + South + Midwest, data = flu\_data2) summary(flu.lm2) #0.7041

vif(flu.lm2) # have to without interactions because automatically creates multicollinearity ols\_test\_breusch\_pagan(flu.lm2) # passes

flu.lm2 <-lm(Flu2~ Income + Uninsured + Religion + Hispanic + Black + Asian + Death2 + Public + Hospitals + Children + Medicaid + Visit + Population55 + Pop75Plus + Mrexemp + Philoexemp + Democratic + College + Cellular + West + South + Midwest, data = flu\_data)

summary(flu.lm2) vif(flu.lm2)

#with geo interactions #FINAL MODEL

flu.lm2 <-lm(Flu~ Income + Visit + Uninsured + Medicaid + Hospitals + Hispanic + Black + Asian + Children + Population55 + Population65Plus + College + Religion + Democratic + Death + Cellular + + Public + Mrexemption + Philoexemption + Philoexemption\*Income + South + Midwest + Midwest + West + Philoexemption\*South + Philoexemption\*Midwest, data = flu\_datafinal)

summary(flu.lm2) ols\_test\_breusch\_pagan(flu.lm2) vif(flu.lm2)

resetTest(flu.lmA, power = 2:3, type = c("fitted")) resetTest(flu.lmA, power = 2, type = c("fitted")) resetTest(flu.lmA, power = 3, type = c("fitted"))

#without interactions for testing multicollinearity

flu.lm2 <-lm(Flu~ Income + Visit + Uninsured + Medicaid + Hospitals + Hispanic + Black + Asian + Children + Population55 + Population65Plus + College + Religion + Democratic + Death + Cellular + + Public + Mrexemption + Philoexemption + South + Midwest + Midwest + West, data = flu\_datafinal) summary(flu.lm2)

ols\_test\_breusch\_pagan(flu.lm2) vif(flu.lm2)

resetTest(flu.lmA, power = 2:3, type = c("fitted")) resetTest(flu.lmA, power = 2, type = c("fitted")) resetTest(flu.lmA, power = 3, type = c("fitted"))

# Works Cited

“2016-2017 Estimated Influenza Illnesses, Medical Visits, Hospitalizations, and Deaths Averted by Vaccination in the United States.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 10 Dec. 2019,

[www.cdc.gov/flu/vaccines-work/burden-averted-2016-17.htm](http://www.cdc.gov/flu/vaccines-work/burden-averted-2016-17.htm).

“About Measles Vaccination.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 28 Mar. 2019, [www.cdc.gov/vaccines/vpd/measles/index.html](http://www.cdc.gov/vaccines/vpd/measles/index.html).

“Adults Who Had a Flu Shot Within the Past Year.” *The Henry J. Kaiser Family Foundation*, 10 Mar.

2020,

[www.kff.org/other/state-indicator/adult-who-had-a-flu-shot-within-the-past-year/?currentTimefra](http://www.kff.org/other/state-indicator/adult-who-had-a-flu-shot-within-the-past-year/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc) [me=0&sortModel={"colId":"Location","sort":"asc](http://www.kff.org/other/state-indicator/adult-who-had-a-flu-shot-within-the-past-year/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)"}.

“Adults Who Report Not Seeing a Doctor in the Past 12 Months Because of Cost by Race/Ethnicity.”

*The Henry J. Kaiser Family Foundation*, 2 Jan. 2020,

[www.kff.org/other/state-indicator/percent-of-adults-reporting-not-seeing-a-doctor-in-the-past-12-](http://www.kff.org/other/state-indicator/percent-of-adults-reporting-not-seeing-a-doctor-in-the-past-12-months-because-of-cost-by-raceethnicity/?currentTimeframe=1&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc) [months-because-of-cost-by-raceethnicity/?currentTimeframe=1&sortModel={"colId":"Location",](http://www.kff.org/other/state-indicator/percent-of-adults-reporting-not-seeing-a-doctor-in-the-past-12-months-because-of-cost-by-raceethnicity/?currentTimeframe=1&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc) ["sort":"asc](http://www.kff.org/other/state-indicator/percent-of-adults-reporting-not-seeing-a-doctor-in-the-past-12-months-because-of-cost-by-raceethnicity/?currentTimeframe=1&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)"}.

“CDC - State School and Childcare Vaccination Laws - Publications by Topic - Public Health Law.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 28 Apr. 2017, [www.cdc.gov/phlp/publications/topic/vaccinations.html](http://www.cdc.gov/phlp/publications/topic/vaccinations.html).

“Disease Burden of Influenza.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 17 Apr. 2020, [www.cdc.gov/flu/about/burden/index.html](http://www.cdc.gov/flu/about/burden/index.html).

“EDUCATIONAL ATTAINMENT .” *Data.census.gov*, 2017,

data.census.gov/cedsci/table?q=education&g=0100000US.04000.001&hidePreview=true&tid=A CSST1Y2017.S1501&t=Education.

“Estimates of Influenza Vaccination Coverage among Adults-United States, 2017–18 Flu Season.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 5 Nov. 2018, [www.cdc.gov/flu/fluvaxview/coverage-1718estimates.htm](http://www.cdc.gov/flu/fluvaxview/coverage-1718estimates.htm).

“Health Insurance Coverage of the Total Population.” *The Henry J. Kaiser Family Foundation*, 19 Dec.

2019,

[www.kff.org/other/state-indicator/total-population/?currentTimeframe=1&sortModel={"colId":"](http://www.kff.org/other/state-indicator/total-population/?currentTimeframe=1&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc) [Location","sort":"asc](http://www.kff.org/other/state-indicator/total-population/?currentTimeframe=1&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)"}.

“History of Vaccine Safety History.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 27 Oct. 2015, [www.cdc.gov/vaccinesafety/ensuringsafety/history/index.html](http://www.cdc.gov/vaccinesafety/ensuringsafety/history/index.html).

“MEDIAN INCOME IN THE PAST 12 MONTHS (IN 2017 INFLATION-ADJUSTED DOLLARS) .”

*Data.census.gov*, Unites States Census Bureau, 2017, data.census.gov/cedsci/table?q=income by state&g=0400000US39,32,31,78,34,33,36,35,38,37,72,30,29,28,21,20,23,22,66,69,25,24,27,26,6 0,18,17,19,54,10,53,56,12,55,11,13,16,15,50,51,06,09,08,42,45,44,47,02,46,05,49,04,48,41,40&h

idePreview=true&tid=ACSST5Y2017.S1903&t=Income (Households, Families, Individuals)&vintage=2018.

“Population Distribution by Age.” *The Henry J. Kaiser Family Foundation*, 4 Dec. 2019, [www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel={"colId](http://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc) [":"Location","sort":"asc](http://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)"}.

“Population Distribution by Age.” *The Henry J. Kaiser Family Foundation*, 4 Dec. 2019, [www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel={"colId](http://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc) [":"Location","sort":"asc](http://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)"}.

“Population Distribution by Family Structure.” *The Henry J. Kaiser Family Foundation*, 4 Dec. 2019, [www.kff.org/other/state-indicator/distribution-by-family-structure/?currentTimeframe=1&sortMo](http://www.kff.org/other/state-indicator/distribution-by-family-structure/?currentTimeframe=1&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc) [del={"colId":"Location","sort":"asc](http://www.kff.org/other/state-indicator/distribution-by-family-structure/?currentTimeframe=1&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)"}.

“Population Distribution by Race/Ethnicity.” *The Henry J. Kaiser Family Foundation*, 4 Dec. 2019, [www.kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=0&sortModel](http://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)

[={"colId":"Location","sort":"asc](http://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=0&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc)"}.

“Stats of the States - Influenza/Pneumonia Mortality.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 19 Feb. 2020, [www.cdc.gov/nchs/pressroom/sosmap/flu\_pneumonia\_mortality/flu\_pneumonia.htm](http://www.cdc.gov/nchs/pressroom/sosmap/flu_pneumonia_mortality/flu_pneumonia.htm).

“Stats of the States - Influenza/Pneumonia Mortality.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 19 Feb. 2020, [www.cdc.gov/nchs/pressroom/sosmap/flu\_pneumonia\_mortality/flu\_pneumonia.htm](http://www.cdc.gov/nchs/pressroom/sosmap/flu_pneumonia_mortality/flu_pneumonia.htm).

“The Flu Season.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 12 July 2018, [www.cdc.gov/flu/about/season/flu-season.htm](http://www.cdc.gov/flu/about/season/flu-season.htm).

“TYPES OF COMPUTERS AND INTERNET SUBSCRIPTIONS .” *Data.census.gov*, Dec. 2017,

data.census.gov/cedsci/table?q=computers&tid=ACSST1Y2017.S2801&vintage=2018. “Vaccine Effectiveness: How Well Do the Flu Vaccines Work?” *Centers for Disease Control and*

*Prevention*, Centers for Disease Control and Prevention, 3 Jan. 2020, [www.cdc.gov/flu/vaccines-work/vaccineeffect.htm](http://www.cdc.gov/flu/vaccines-work/vaccineeffect.htm).

Benecke, Olivia, and Sarah Elizabeth DeYoung. “Anti-Vaccine Decision-Making and Measles Resurgence in the United States.” *Global Pediatric Health*, SAGE Publications, 24 July 2019, [www.ncbi.nlm.nih.gov/pmc/articles/PMC6657116/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC6657116/).

Duffin, Erin. “Births and Deaths in the U.S. by State 2017.” *Statista*, 1 July 2019, [www.statista.com/statistics/241581/births-and-deaths-in-the-us-by-state/](http://www.statista.com/statistics/241581/births-and-deaths-in-the-us-by-state/).

Gallup. “2017 U.S. Party Affiliation by State.” *Gallup.com*, Gallup, 2 Feb. 2018, news.gallup.com/poll/226643/2017-party-affiliation-state.aspx.

Lipka, Michael, and Benjamin Wormald. “Most and Least Religious U.S. States.” *Pew Research Center*, Pew Research Center, 29 Feb. 2016,

[www.pewresearch.org/fact-tank/2016/02/29/how-religious-is-your-state/?state=alabam](http://www.pewresearch.org/fact-tank/2016/02/29/how-religious-is-your-state/?state=alabam).

Patel, Manisha, et al. “Increase in Measles Cases — United States, January 1–April 26, 2019.” *Morbidity and Mortality Weekly Report*, CDC, 3 May 2019, [www.cdc.gov/mmwr/volumes/68/wr/pdfs/mm6817e1-H.pdf](http://www.cdc.gov/mmwr/volumes/68/wr/pdfs/mm6817e1-H.pdf).

Putri, Wayan C W S, et al. “Economic Burden of Seasonal Influenza in the United States.” *Vaccine*, U.S. National Library of Medicine, 22 June 2018, [www.ncbi.nlm.nih.gov/pubmed/29801998](http://www.ncbi.nlm.nih.gov/pubmed/29801998).

Santibanez, Tammy A, et al. “Socio-Demographic Differences in Opinions about 2009 Pandemic Influenza A (H1N1) and Seasonal Influenza Vaccination and Disease among Adults during the 2009-2010 Influenza Season.” *Influenza and Other Respiratory Viruses*, Blackwell Publishing Ltd, May 2013, [www.ncbi.nlm.nih.gov/pubmed/22568588](http://www.ncbi.nlm.nih.gov/pubmed/22568588).

Siddiqui, Mariam, et al. “Epidemiology of Vaccine Hesitancy in the United States.” *Human Vaccines & Immunotherapeutics*, Landes Bioscience, Dec. 2013, [www.ncbi.nlm.nih.gov/pubmed/24247148](http://www.ncbi.nlm.nih.gov/pubmed/24247148).

Stern, Alexandra Minna, et al. “The History Of Vaccines And Immunization: Familiar Patterns, New Challenges.” Health Affairs, June 2005, [www.healthaffairs.org/doi/full/10.1377/hlthaff.24.3.611](http://www.healthaffairs.org/doi/full/10.1377/hlthaff.24.3.611).