What is Most Important in Determining Heart Disease & Stroke?

Brittany Burlison

Oklahoma State University

brittany.burlison@okstate.edu

Kailey Wilson

Oklahoma State University

kailey.r.wilson@okstate.edu

Abstract:

Heart disease and strokes are two major diseases that have been around for years and there still is no cure for. Heart disease is the leading cause of death in the United States. A person dies every 36 seconds from heart disease. Of these deaths, 1 in 6 people die due to a stroke. Strokes are the leading cause for long-term disabilities. For our research project, we will look and see if these two major diseases have common factors that will be able to predict each other. First, we will build a logistic regression model for each disease. Then, we will make a new variable that returns 1 if the person has both diseases and 0 if not. Finally, we will do a final analysis to see which variables in these two models will be able to predict both diseases in one equation. From our research we will find that the variables general health, diabetes and health coverage will do the best at helping to determine whether or not a person will suffer from heart disease or a stroke in their lifetime.

Introduction:

For our research project we are interested in seeing if there are factors that are more important in determining whether a person will suffer from a stroke or heart disease in their lifetime. We are wanting

to take variables that correlate to the Social Determinants of Health and see what variables play a bigger role in determining these major health issues.

Data Overview:

The data we will be using for our analysis is from the Behavioral Risk Factor Surveillance System (BRFSS) from the CDC. This is a phone survey that collects data from citizens regarding a plethora of information. We will be using data from 2016-2020. This contains over 500 fields and over 2 million observations. Some of the fields contain information about the household, current health conditions, behaviors, and demographics. Additionally, some states have the option to have more specific health questions to be considered. We will be looking at variables that all people are asked.

Method and Plans:

Our data set contains over 500 variables so we have narrowed that list down to 11 that we have deemed the most important in determining heart disease or stroke. We referenced the social determinants of health, which can be found in the appendix under Table 1, to help us make the decision on which variables we should keep. We have determined that Income, Housing, Education, Mental Health, Health Coverage, Overall General Health, Smoking status, Diabetes, State, Divorce, and Medical Cost were the most important variables to look at. The list of SAS Variables used can be found in the Appendix under Table 2. We will be using Stroke and Heart Disease as our response variables. We will look at these variables by Gender using the LANDSEX variable. We will then concatenate all 5 years of our data in JMP and run a fit model test to determine which preselected variables are the most important in determining heart disease or stroke.

Results:

From our analysis we have drawn the following conclusions:

We found that the most important variables in determining whether or not a person will suffer from a stroke is diabetes, general health, & education for females. While for males, the variables are diabetes, general health, and health insurance. We can determine this by looking at the p-value for the variables stated. Screenshots of the results for both males and females can be found in the appendix under Image 1 and Image 2.

We found that the most important variables in determining whether or not a person will have heart disease is general health, diabetes, smoking, & if their parents are divorced for males. While for females, it is general health, diabetes, smoking, & income. We can determine this by looking at the p-value for the variables tested. Screenshots of the results for both males and females can be found in the appendix under Image 3 and 4.

We found that to determine if a person will suffer from both heart disease & stroke, the most important variables in determining this are general health, diabetes, income, if they smoke and their education.

We found that the most important variables in determining whether or not a person will suffer from a stroke is diabetes, general health, & education for females. While for males, the variables are diabetes, general health, and health insurance. We determined these results by looking at the p-value for our variables. Screenshots for these results can be found in the Appendix under Image5. Since general health was a major variable in determining heart disease and stroke we drilled down and graphed what each response of general health would look like in determining heart disease and/or stroke. We found that across the board that response 2, meaning very good overall health, and 3, meaning good overall health, had the highest chance of either suffering from heart disease and/or stroke. These graphs can be found in the Appendix under Image 6, 7, and 8.

Implications:

Overall these are the major implications we found from our conclusions of our research:

1. To help prevent heart disease, people should improve their general health, monitor their diabetes, and decrease their nicotine use.

2. To help prevent stroke, people should improve their general health, monitor their diabetes, and think about improving their health care plan.
3. Overall, people should focus on their general health to prevent heart disease and stroke.

We believe that if doctors and health care providers take into consideration these important factors in determining whether a person will suffer from heart disease or stroke in their lifetime they will be able to provide better health care to their patients. Additionally, we feel that if the general public takes these factors into consideration it can help reduce the risk of stroke or heart disease.

Citations

CDC – National Center for Health Statistics – 2016 BRFSS Survey Data and Documentation

https://www.cdc.gov/brfss/annual_data/annual_2016.html

CDC – National Center for Health Statistics – 2017 BRFSS Survey Data and Documentation

https://www.cdc.gov/brfss/annual_data/annual_2017.html

CDC – National Center for Health Statistics – 2018 BRFSS Survey Data and Documentation

https://www.cdc.gov/brfss/annual_data/annual_2018.html

CDC – National Center for Health Statistics – 2019 BRFSS Survey Data and Documentation

https://www.cdc.gov/brfss/annual_data/annual_2019.html

CDC – National Center for Health Statistics – 2020 BRFSS Survey Data and Documentation

https://www.cdc.gov/brfss/annual_data/annual_2020.html

APPENDIX

Table 1

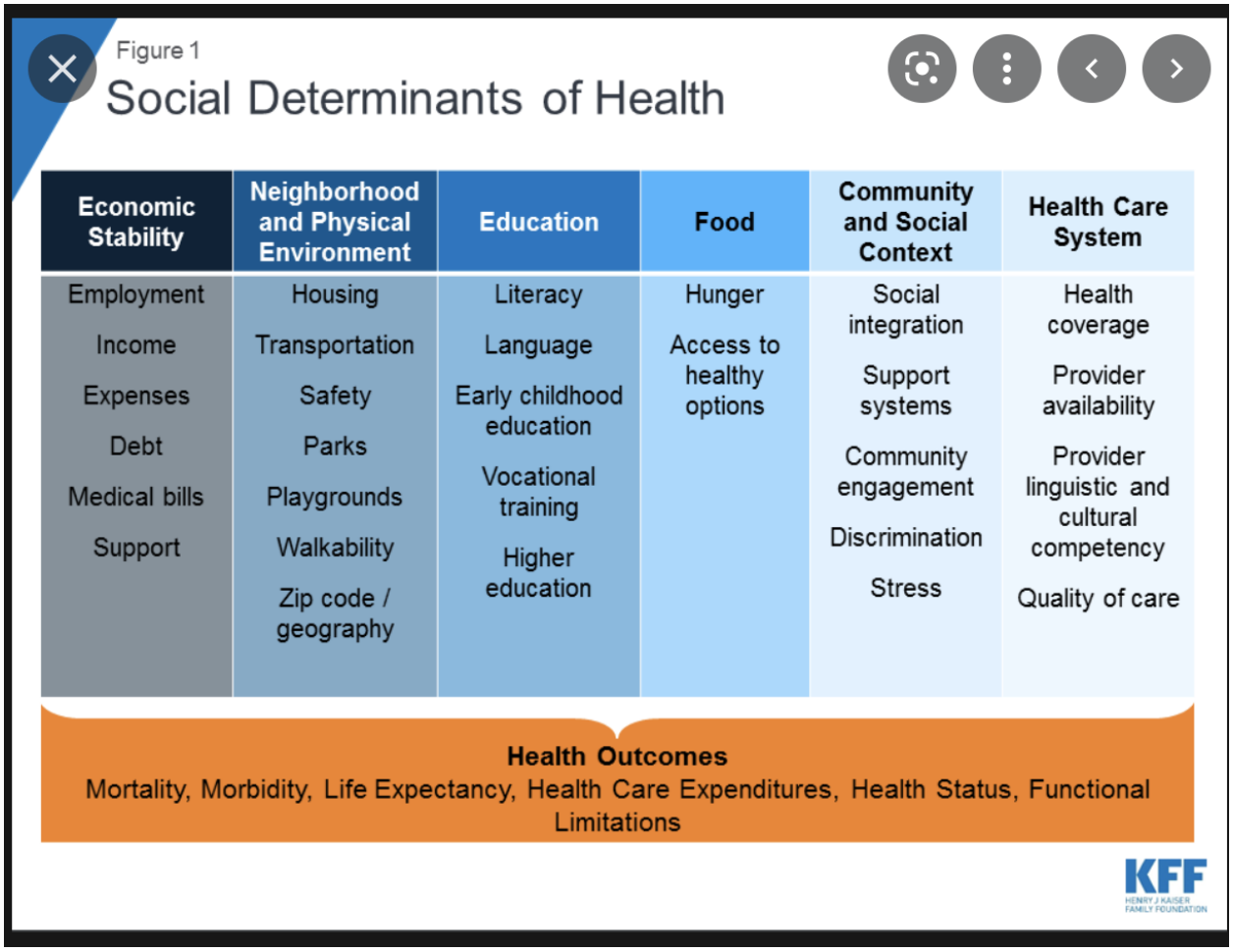
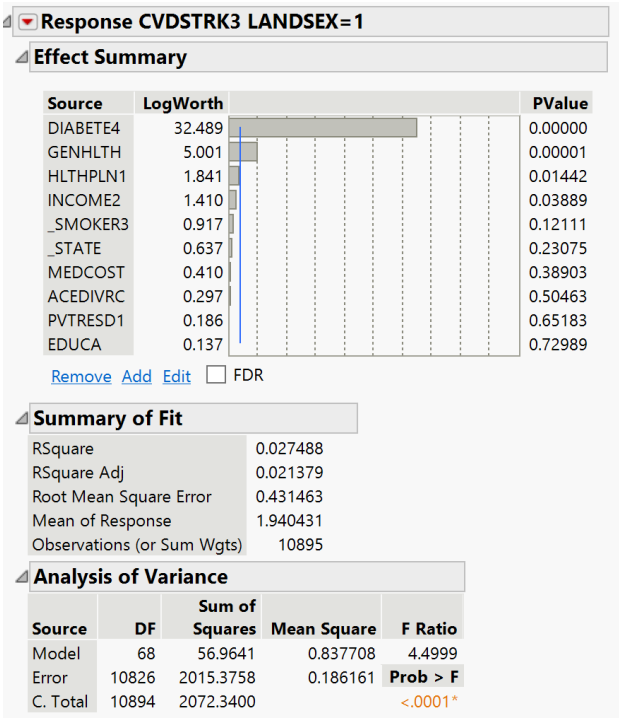


Table 2

Variable Name
INCOME2
PVTRESD1
EDUCA
MNTHLTH
HLTHPLN1
GENHLTH
ACEDIVRC
_SMOKER3
DIABETE4
MEDCOST
_STATE

Image 1 (Stroke - Male)



Response CVDSTRK3 LANDSEX=1

Effect Summary

Source	LogWorth	PValue
DIABETE4	32.489	0.00000
GENHLTH	5.001	0.00001
HLTHPLN1	1.841	0.01442
INCOME2	1.410	0.03889
_SMOKER3	0.917	0.12111
_STATE	0.637	0.23075
MEDCOST	0.410	0.38903
ACEDIVRC	0.297	0.50463
PVTRESD1	0.186	0.65183
EDUCA	0.137	0.72989

Remove Add Edit ☐ FDR

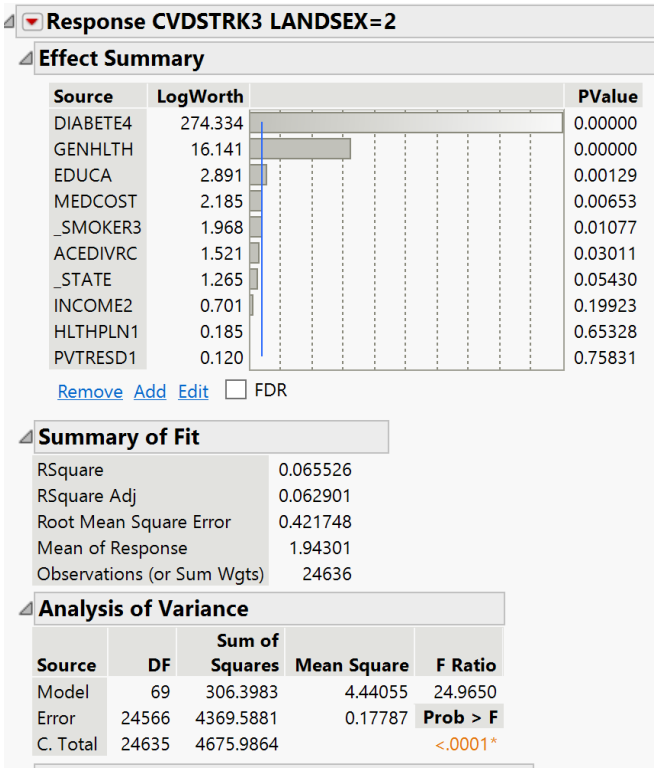
Summary of Fit

RSquare	0.027488
RSquare Adj	0.021379
Root Mean Square Error	0.431463
Mean of Response	1.940431
Observations (or Sum Wgts)	10895

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	68	56.9641	0.837708	4.4999
Error	10826	2015.3758	0.186161	Prob > F
C. Total	10894	2072.3400		<.0001*

Image 2 (Stroke - Female)



Response CVDSTRK3 LANDSEX=2

Effect Summary

Source	LogWorth	PValue
DIABETE4	274.334	0.00000
GENHLTH	16.141	0.00000
EDUCA	2.891	0.00129
MEDCOST	2.185	0.00653
_SMOKER3	1.968	0.01077
ACEDIVRC	1.521	0.03011
_STATE	1.265	0.05430
INCOME2	0.701	0.19923
HLTHPLN1	0.185	0.65328
PVTRESD1	0.120	0.75831

Remove Add Edit ☐ FDR

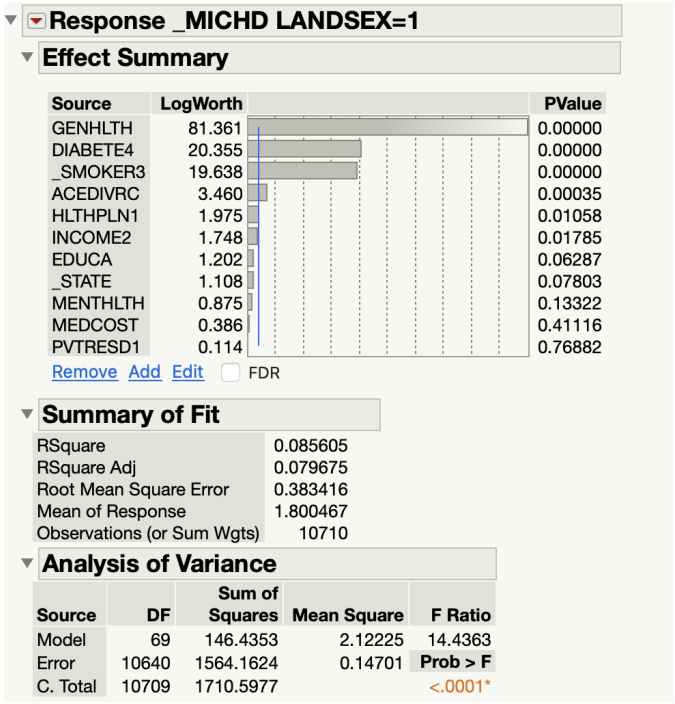
Summary of Fit

RSquare	0.065526
RSquare Adj	0.062901
Root Mean Square Error	0.421748
Mean of Response	1.94301
Observations (or Sum Wgts)	24636

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	69	306.3983	4.44055	24.9650
Error	24566	4369.5881	0.17787	Prob > F
C. Total	24635	4675.9864		<.0001*

Image 3 (Heart Disease - Male)



Response _MICHD LANDSEX=1

Effect Summary

Source	LogWorth	PValue
GENHLTH	81.361	0.00000
DIABETE4	20.355	0.00000
_SMOKER3	19.638	0.00000
ACEDIVRC	3.460	0.00035
HLTHPLN1	1.975	0.01058
INCOME2	1.748	0.01785
EDUCA	1.202	0.06287
_STATE	1.108	0.07803
MENTHLTH	0.875	0.13322
MEDCOST	0.386	0.41116
PVTRESD1	0.114	0.76882

Remove Add Edit FDR

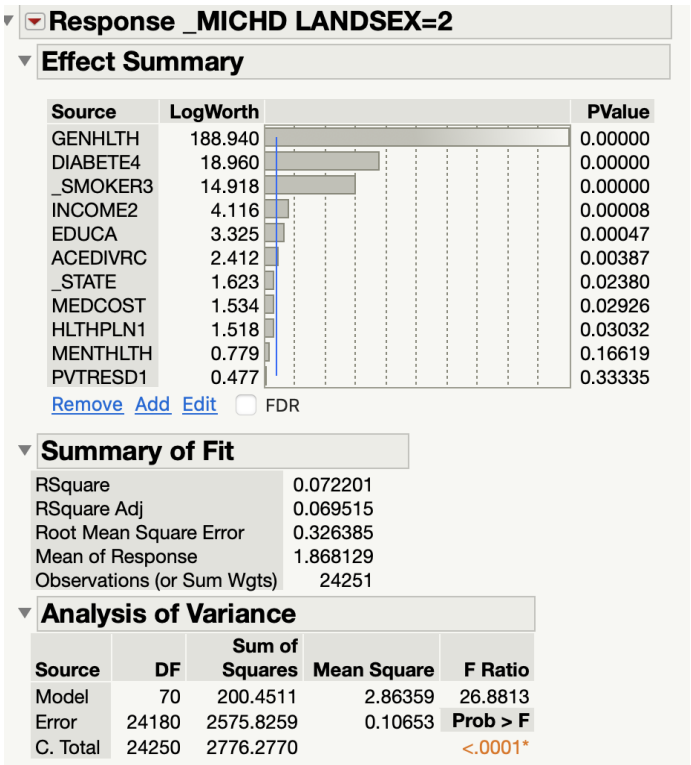
Summary of Fit

RSquare	0.085605
RSquare Adj	0.079675
Root Mean Square Error	0.383416
Mean of Response	1.800467
Observations (or Sum Wgts)	10710

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	69	146.4353	2.12225	14.4363
Error	10640	1564.1624	0.14701	Prob > F
C. Total	10709	1710.5977		<.0001*

Image 4 (Heart Disease - Female)



Response _MICHD LANDSEX=2

Effect Summary

Source	LogWorth	PValue
GENHLTH	188.940	0.00000
DIABETE4	18.960	0.00000
_SMOKER3	14.918	0.00000
INCOME2	4.116	0.00008
EDUCA	3.325	0.00047
ACEDIVRC	2.412	0.00387
_STATE	1.623	0.02380
MEDCOST	1.534	0.02926
HLTHPLN1	1.518	0.03032
MENTHLTH	0.779	0.16619
PVTRESD1	0.477	0.33335

Remove Add Edit FDR

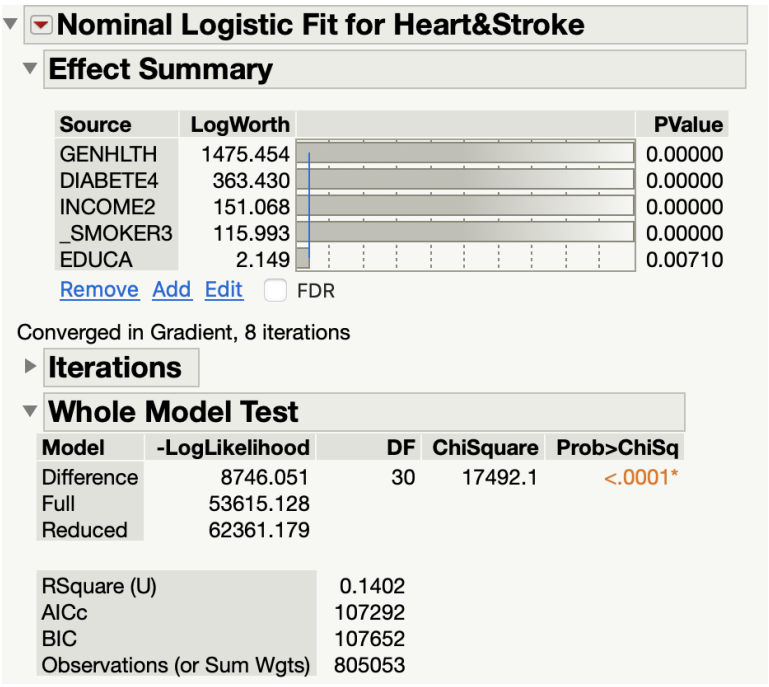
Summary of Fit

RSquare	0.072201
RSquare Adj	0.069515
Root Mean Square Error	0.326385
Mean of Response	1.868129
Observations (or Sum Wgts)	24251

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	70	200.4511	2.86359	26.8813
Error	24180	2575.8259	0.10653	Prob > F
C. Total	24250	2776.2770		<.0001*

Image 5 (Heart Disease & Stroke)



Nominal Logistic Fit for Heart&Stroke

Effect Summary

Source	LogWorth	PValue
GENHLTH	1475.454	0.00000
DIABETE4	363.430	0.00000
INCOME2	151.068	0.00000
_SMOKER3	115.993	0.00000
EDUCA	2.149	0.00710

Remove Add Edit FDR

Converged in Gradient, 8 iterations

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	8746.051	30	17492.1	<.0001*
Full	53615.128			
Reduced	62361.179			

RSquare (U)	0.1402
AICc	107292
BIC	107652
Observations (or Sum Wgts)	805053

Image 6 (Heart Disease - Graph)

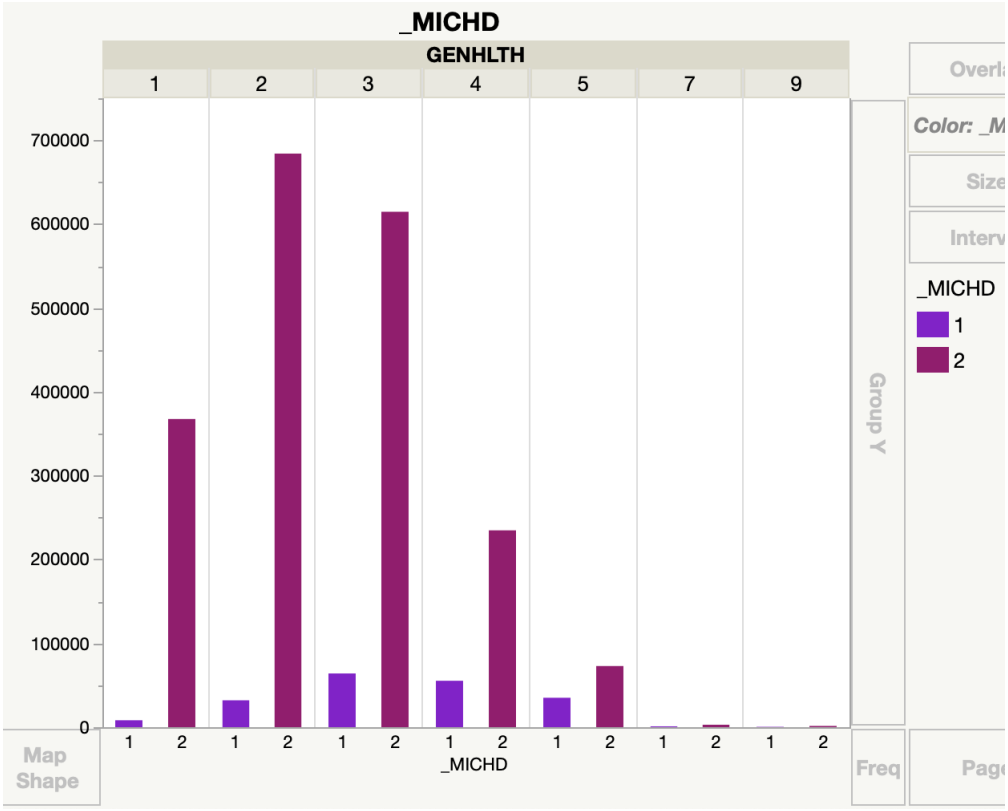


Image 7 (Stroke - Graph)

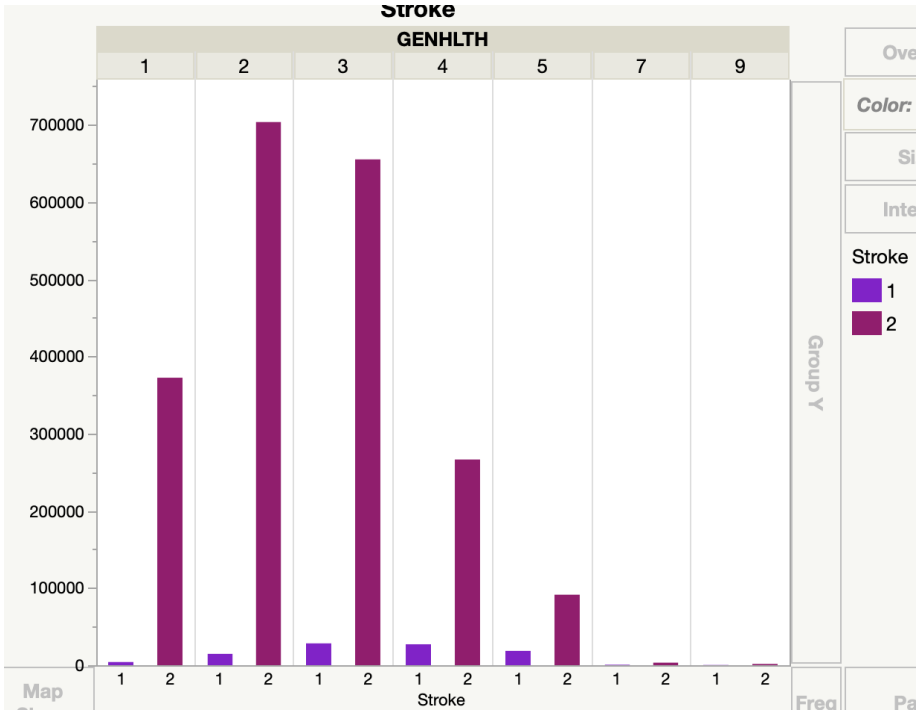


Image 8 (Heart Disease & Stroke - Graph)