Are Wealthier People Healthier ?

Amrita Veerabagu, Devarsh Jhaveri, Neehar Yalamarti and Siddharth Ajit

The relationship between the factors of health and income is a widely debated and discussed topic today. The higher a person's income, higher is the chance of that person being less healthy and premature death. Improving the income of a person in a family is one of the ways that is proposed to improve their health. In order to substantiate this claim, proper research is needed that shows if there exists a definitive relation between money and health.

According to the "Monthly Labor Review: Income and Health Outcomes", Serah Hyde analyzes that the life expectancy of the average American, which has increased substantially over the past 50 years. Although this is a good sign, a close examination of the health indicators reveals troubling signs. The obesity and diabetes rates have increased manifold between 2008 and 2013 as is also shown below.

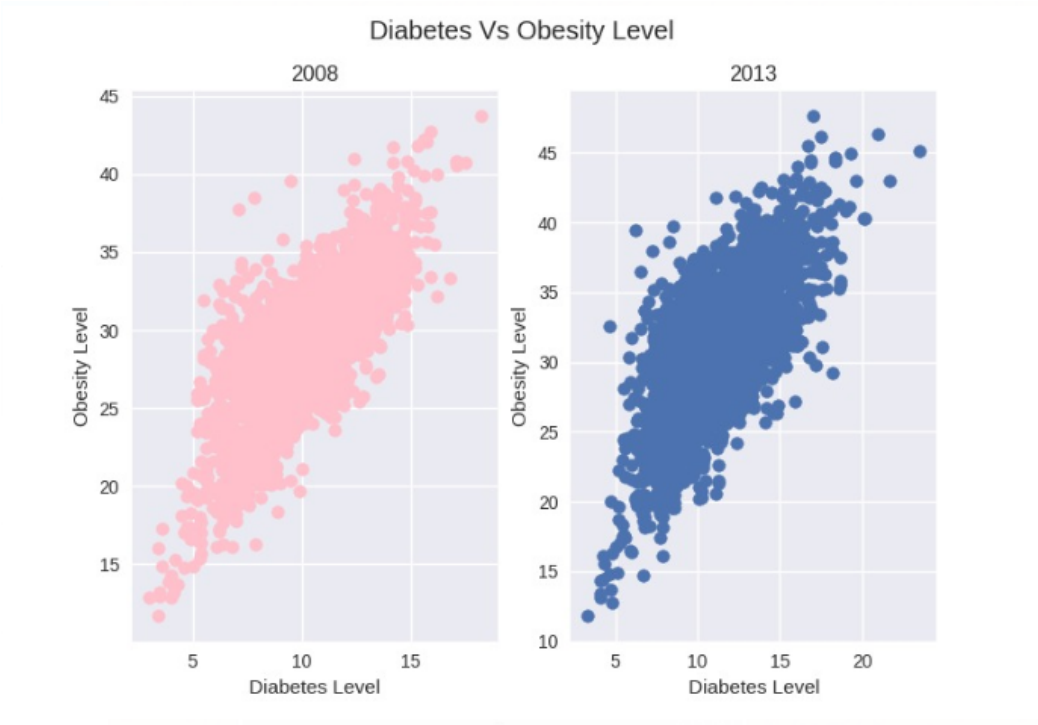

Fig 1: Plot between obesity and diabetes for the year 2008 and 2013

Analyzing how diabetes and obesity values are distributed in 2008 and 2013, it can be observed that the upper limit values of both diabetes and obesity have increased from 2008 to 2013. In addition, there's a general linear relationship between the two variables for both the years.

Observing the trend state wise, Arkansas has a major increase in obesity rates from 2008 to 2013, closely followed by Iowa, both increasing by approximately 15%. Only three states: California, Nevada and Hawaii have had a decrease in obesity rate whereas the rest of the states have shown an increase, affirming that health is in a general state of decline in US.

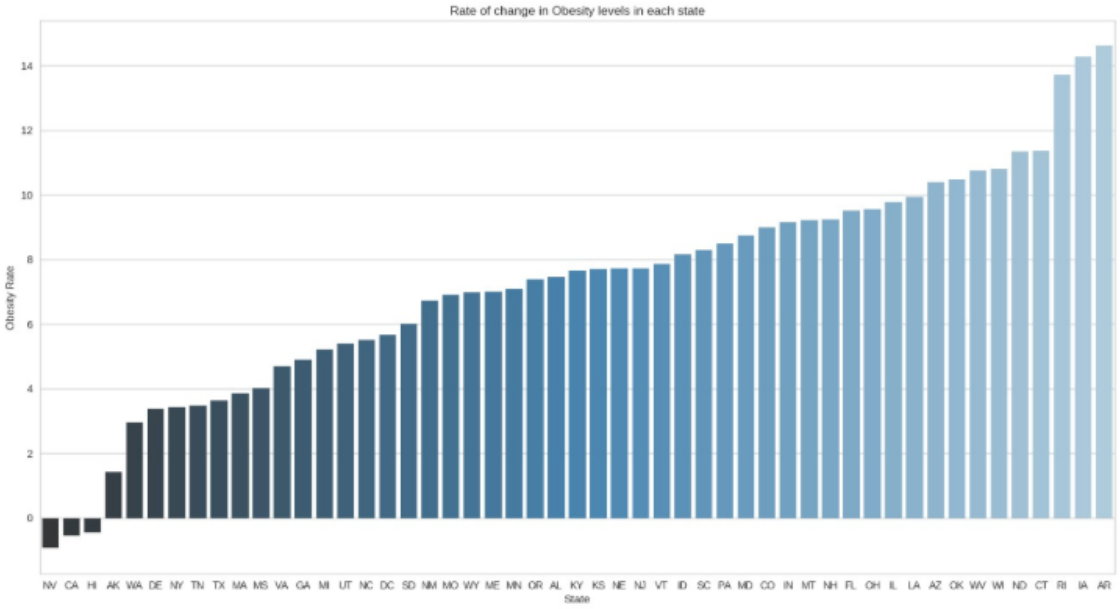

Fig 2: Rate of change of Obesity levels from 2008 to 2013 per state.

Similar inferences can also be drawn from the diabetes rates between 2008 and 2013. Apart from District of Colombia, all states have had an increase in diabetes rates with New Mexico leading the pack with approximately 30%. In addition, diabetes is also expensive to treat, bringing in income as a factor. The median income per person in New Mexico is low, implying insufficient income to treat diabetes problems.

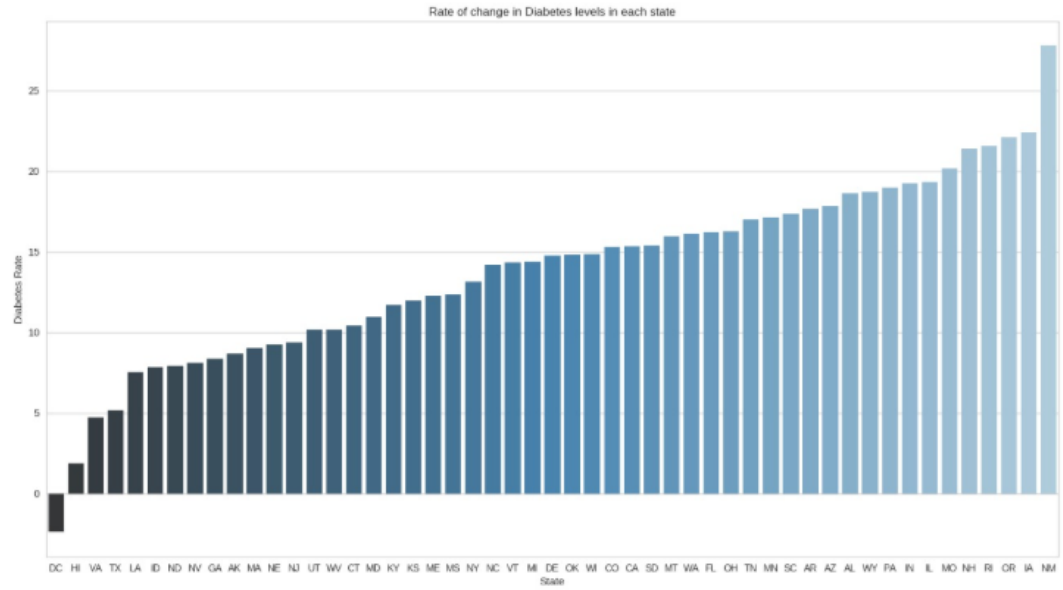

Fig 3: Rate of change of Diabetes levels from 2008 to 2013 per state.

A better visualization can be made state wise for Obesity for the year 2013:

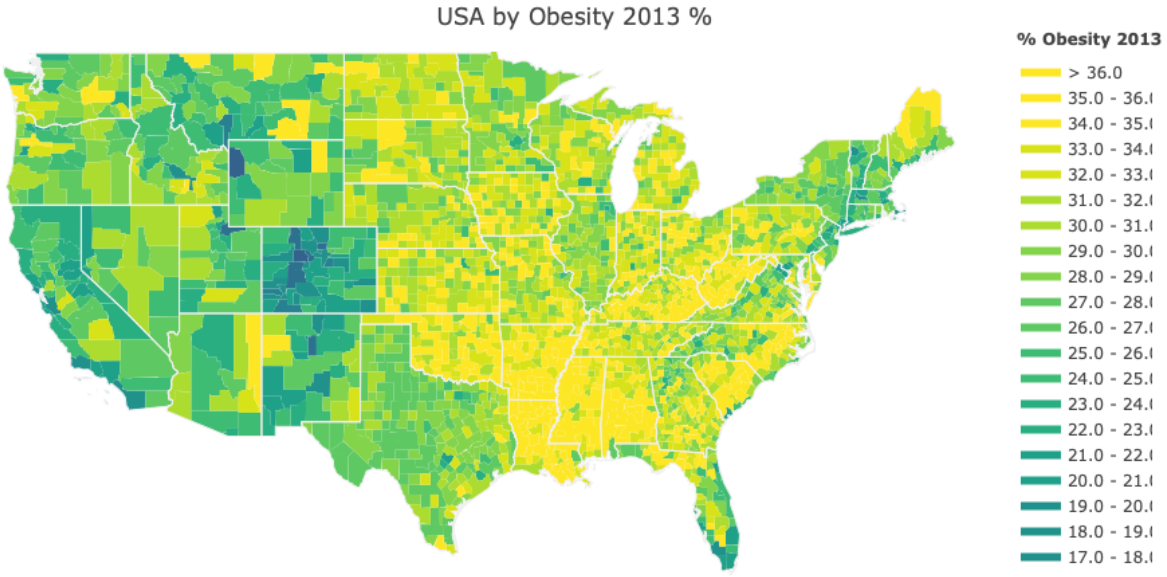

Fig 4: Obesity levels by county in 2013.

In the Southern US States i.e. Virginia, Tennessee, North Carolina, South Carolina, Alabama, Florida and Texas, the obesity levels are on the higher end of the spectrum. In addition, the states of Kansas, Nebraska, Iowa, Missouri, West Virginia, Maine and Ohio, a greater increase in level of obesity has been witnessed. Colorado is by far the best of all having the lowest levels of Obesity in the entire United States of America.

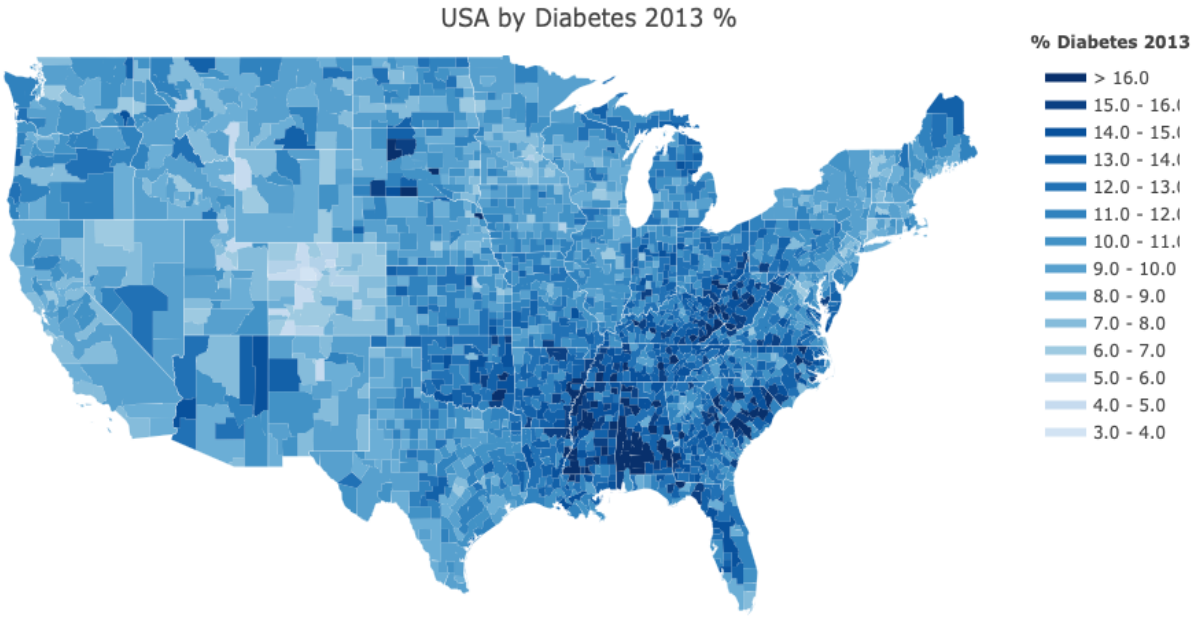

Fig 5: Diabetes levels by county in 2013.

Analyzing the percentage of diabetes in all US Counties in 2013, the Southern part of the US like Alabama, Tennessee, Arkansas, Louisiana and- South Carolina, have witnessed a higher rate in

the diabetes as compared to the rest. The Western and Northern part of the USA are on the lower end of the spectrum of diabetes level. Similarly, from the observations of the pervious plot, we can see that the state of Colorado is by far the best, having the lowest level of diabetes.

From the obesity and diabetes plots, it can be inferred that there is a similarity in the distribution of the respective factors over the entire US. It's not surprising to see this relation. Studies such as "The epidemic of obesity and diabetes, Texas Heart Institute Journal" have shown that Obesity and Diabetes are highly correlated.

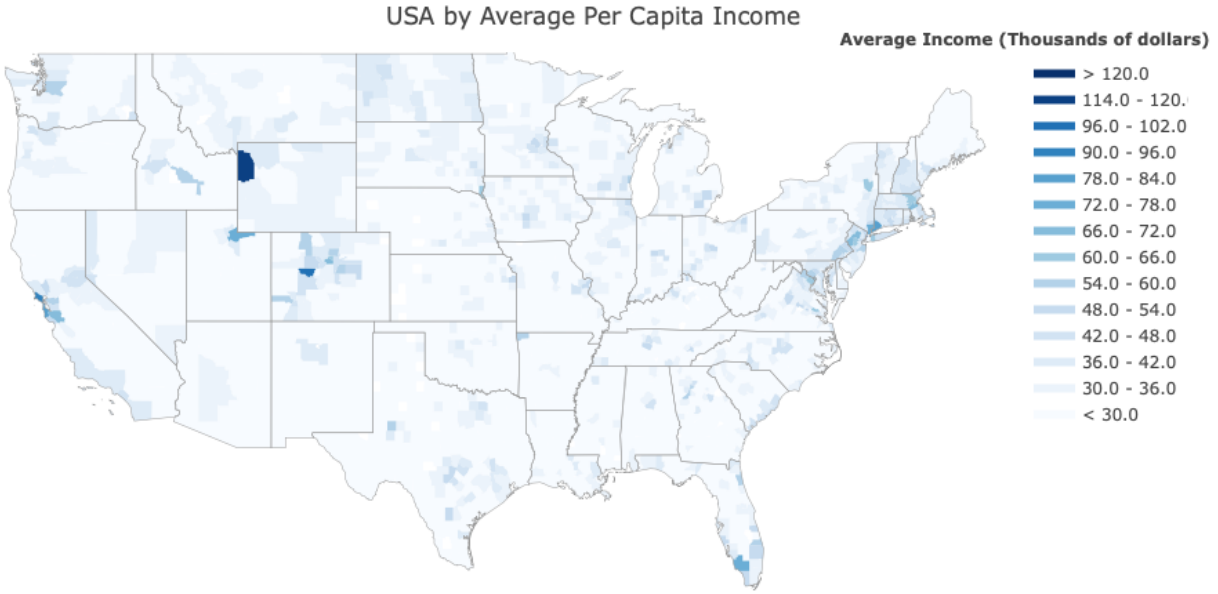

Fig 6: Average AGI per county in the year 2016.

When it comes to analyzing the distribution of income by county, most of the counties in the US are on the lower end i.e. between $30,000 and $50,000. It is also observed that only minimal number of counties have a very high average AGI level as compared to the entire USA, indicating the higher wage gap between income strata in society. In "Money lightens the load", Diane Whitmore notes that top earners have experienced constant income growth, while middle and low-income groups have experienced the opposite, weak growth. This information provides a motivation for further investigation of the differences of health indicators in their respective income strata's.

The obesity and diabetes levels are considered as two of the many indicators of health which are also aligned with the average per capita income in the United States of America. The wealthiest county as seen from the plot is Teton, Wyoming, which as seen from previous plots, have obesity and diabetes levels on the lower end of the spectrum. Thus, an inference can be made on the correlation between income and health. However, a general conclusion regarding the correlation between income and health cannot be made as this minimalistic population does not represent the overall condition in the United States of America.

Data?

To overcome the shortcomings of the previous findings, analysis must be done considering all the counties and states to make a generic statement regarding the correlation. First steps towards such analysis is aggregating several county and state wise data about health and income. The dataset used for analysis was prepared from three publicly available datasets: individual income tax returns published by the Internal Revenue Services (IRS) for the year 2016, County health rankings data published by countyhealthrankings.org and ZIP to FIP crosswalk file found on HUD.org.

Aggregated Gross Income (AGI) represents the cumulative sum of incomes for a given zip code. However, a larger AGI does not necessarily indicate a higher income per person.

A new feature of Average Income per person(AIP) was constructed by dividing the adjusted gross income by the number of exemptions (N2). According to the information found on IRS website, the variable N2 is a fair approximation of the population count. So, the variable AIP gives us average income of a population for the respective income slab for all the zip codes.

To find an income value that adequately represents each zip code, average of the variable AIP weighted by the population count (N2) for each income level was adopted. The weighted mean method considers the distribution of income levels, giving a more realistic estimator of population income for a zip code location. Similarly, average income per person for a county was derived by taking a weighted average of income by the population for zip codes in a county.

In the County Health Data, various health and socio-economic indicators are included in the data, from which only the relevant health metrics were extracted. In alignment with our objective, an estimator of income level of population (AIP) aggregated by county, from the Income tax returns data, is compared with the health indicators from the County health rankings data set.

The Income – Health Connection!

The figure below represents the correlation value of health indicators with the average income per person for a county. A correlation is obtained between several indicators of health with income from the dataset which was prepared from the publicly available data on health indicators for each county. Even though richer people having better health is an easier statement to believe, the relationship other way around i.e. determining the association of health to income cannot be ignored. Thus, giving birth to a classic chicken – egg problem!

In addition to this, there cannot be a sole outcome indicating whether a person is healthy or not, as there is no specific health index but just health indicators. For indicators of wealth of a person,

access to fitness facilities or spas etc. are not necessarily indicators but just other ways of capturing the income.

Some of the indicators such as adult obesity value, diabetes value, premature age adjusted death value have high negative correlation with the average income value. These values indicate the percentage of population in the county that have diabetes/obesity.

Inferring from the correlation diagram below, when the average income per county increases, the percentage value of obese adults in a county decrease. Similarly, when the percentage value of obese adults in a county increases, average income per county decreases.

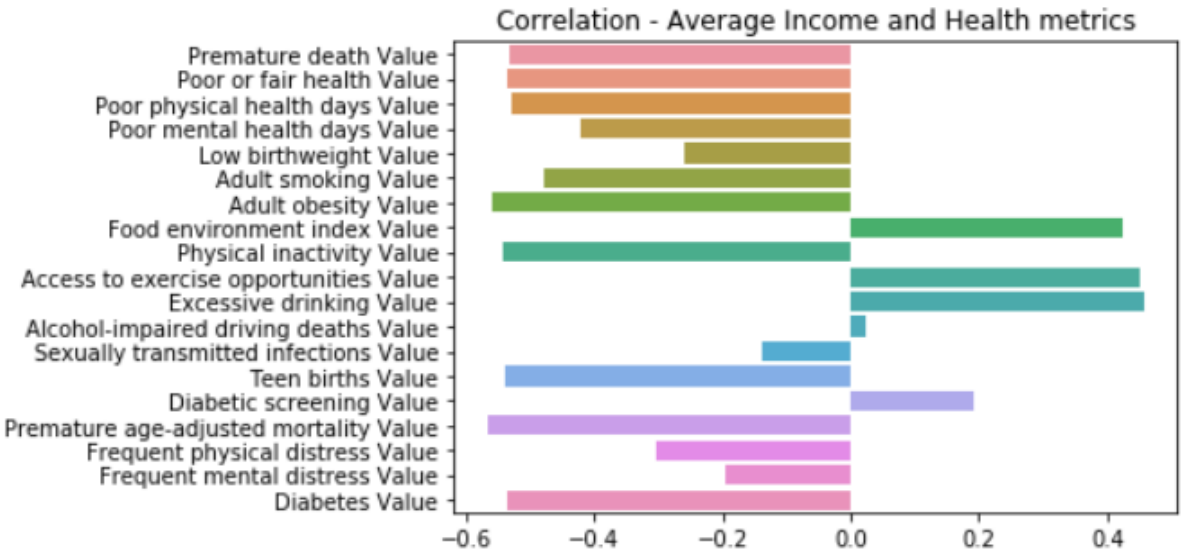

Fig 7: Correlation between the average income and health metrics.

Among the variables that are positively correlated with average income per county, one variable is the Food environment index value indicated on a scale from 1 to 10, indicating whether a person has access to healthy food, with 10 being greater access to healthy food. Healthier food habits, typically result in better well-being which as seen above are positively correlated with average income. The same goes to access to exercise opportunities value and others.

However, a univariate analysis taking a single variable at a time is not appropriate. Hence, we employ advanced statistical algorithms to identify the relationship considering many factors at a time.

An algorithm like the Linear Regression serves as a starting point in identifying the relationship between income and health, considering all the factors. This model results in 50% of the variation in the average income being explained by the health indicators, i.e 50% of the change in income can be attributed to values of health indicators.

To further understand which of these indicators play a major contribution compared to others, another advanced algorithm like the Random Forest can be used. The graph below shows the feature (variable) importance as compared to others, a higher value indicating greater importance in explaining the outcome i.e. the average income per county.

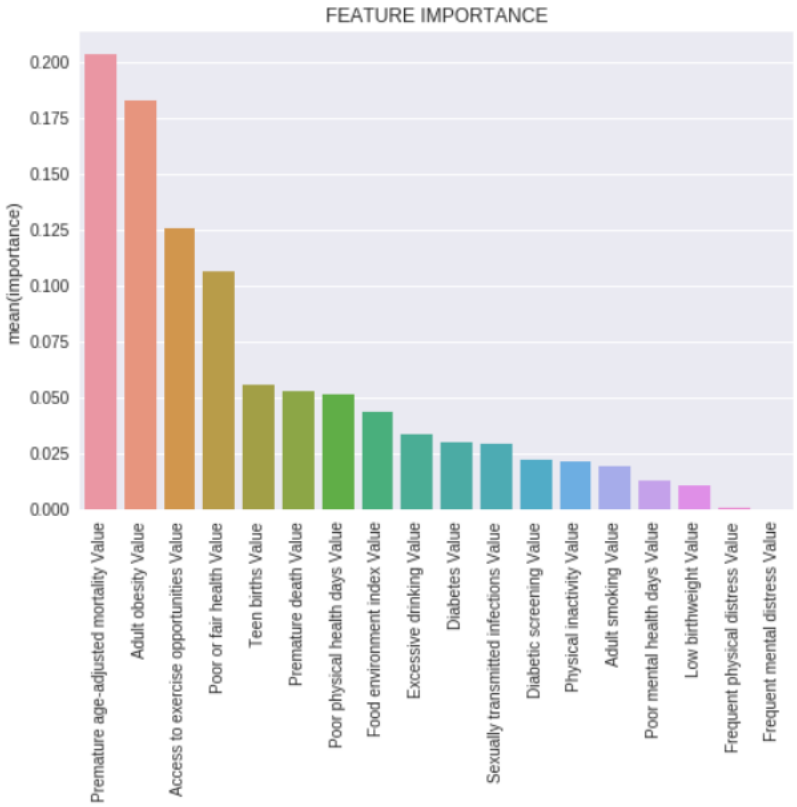

Fig 8: Feature importance w.r.t the average AGI.

Connecting the Dots...

In conclusion, there is a correlation between income and health as seen from the analysis above, but according to the factors mentioned above and the socio-economic factors such as household, race, gender, geography and the disparities in the income level among communities, it cannot be said definitively that higher income is the sole cause of better health and vice-versa. There is reasonable evidence to substantiate the claim that income level is positively correlated to health or negatively correlated to factors indicating depreciating health (example: diabetes percentage value, obesity percentage value)

It is acceptable to believe that people with bad eating habits are more susceptible to poor health. For example, as the processed food like the burgers is very cheap compared to the healthier counterparts, but the rate of consumption of such unhealthy foods cannot be solely attributed to

the fact that people are not economically sound. Income is just one factor that may be dependent on various socio-economic factors resulting in multiple inter-connected pathways. These habits may also arise due to lack of education and awareness.

Health is correlated to income but improving income factors doesn't solely improve health. These could be proxies for some other underlying socio-economic factors. For example, we know that there are more Shark attacks when more ice cream is sold but this doesn't imply that more ice creams being sold solely determines the high frequency of shark attacks. Ice cream being sold is confounding many other variables such as the summer heat or the number of people visiting the beach which could in turn result in the higher frequency of shark attacks.

Similarly, there may be many more confounding effects that indirectly affect health apart from just income and vice versa. Further research is recommended involving the vast array of socio-economic factors and their effect on the respective health indicators in the society. The fact that income and health are correlated is a well-established paradigm, and this merely adds to this claim, however there are multiple inter-connected pathways that need to be investigated.

Bottom Line: *According to the analysis above, it is found that there is a positive correlation between higher income and good health, however, drawing a cause and effect relationship is not straightforward.*