Project 1 Report
Determining Correlations Between United States' Health and Income Data

ECEN 689: Applied Information Science Practicum - Team 7

Course Instructor: Prof. Jean-Francois Chamberland

Team 7 Members

1) Khaled Nakhleh
2) Shabarish Prasad
3) Fazia Batool
4) Venkata Pydimarri

Due Date: October 30th, 2018

Abstract

The goal of this project is to support the hypothesis ”Income level is positively correlated to health”. We analyzed two health indicators, diabetes and obesity rates for the years 2008 and 2013. County income data was analyzed only for the year 2013, since the 2008 income dataset was offered per state, and not for the United States as a whole.

Introduction

Project 1 for ECEN 689 course involved gathering health and income data for the United States. The project also involved parsing through the data, and ensuring accurate representation of U.S. counties. The project’s main objective was to determine if health parameters (obesity, diabetes) were positively correlated with income data collected (adjusted gross income A.G.I.) by the I.R.S. (Internal Revenue Service).

The project goal was also to better understand the relationship between health measurements within the broader US social classes, and the classes’ income disparity proven in previous academic work by Bryan and Martinez [4].

Going through the literature, one can see an established truth about income and health: the higher an individual's income, the better his or her health [1]. However, a broader scope is needed when considering how those parameters are measured county-wise. Not to mention the perceived health evaluation individuals have about their life. We highlight the importance of rationally choosing measurement variables for county-level evaluation. Previous work recommends the variables should be dependent on the relevance of the variable to the model being studied, as L Shavers published [5]. We want to also acknowledge other factors that we believe affect a county’s overall health. Education level was also seen to affect health quality in US social classes by Marmot [6]. More discussion is offered in the results and conclusion sections.

Life evaluation refers to the thoughts that people have about their life when they think about it [2]. While the literature did prove a correlation between perceived mental health and income on an individual level, it also proved no tangible progress beyond an annual income of ~$75,000 [2]. Mental health analysis and observations were not included in this work due to time constraints. We recommend future work to incorporate mental health evaluation into their findings.

The health dataset was downloaded from the U.S. department of agriculture website. The health data includes both parameters from the years 2008 and 2013. The income data was downloaded from the IRS website. County income data was used, and not the provided individual income data. A combined health and data file was generated from the previous downloaded files. All US counties were represented in the generated combined file, except for Hawaii’s Kalawao county. Kalawao county did not have income data in the income file.

We highlight how previous work showed that it cannot be decisively conclude how income inequality affects individual health from population-level studies by Wagstaff and Doorslaer [3]. Our used data is classified as a population-level data, hence, the results here cannot firmly prove a relationship between social class income inequality and health levels. The next sections of this report provide graphs, data visualization, and sources that give better understanding of the health-income relationship in U.S. counties. References are provided at the end of this report.

Data Sets

We used two datasets in this project.

The first dataset is "**Food Environment Atlas".** Food environment factors interact to influence food choices and diet quality. The objectives of the Food Atlas, made available by the U.S. Department of Agriculture, are to assemble statistics on food environment indicators and to provide a spatial overview of access to healthy food. The atlas contains health and well-being indicators such as diabetes and obesity rates.
The current version of the Food Environment Atlas has over 275 variables, including new indicators on access and proximity to a grocery store for sub populations; an indicator on the SNAP Combined Application Project for recipients of Supplemental Security Income (at the State level); and indicators on farmers' markets that report accepting credit cards or report selling baked and prepared food products. Data was used for the years 2008 and 2013.

The second dataset is "**County Income Tax Statistics".** The Internal Revenue Service (IRS) is the revenue service of the United States federal government. It is responsible for collecting taxes and administering the Internal Revenue Code. Public information includes individual tax statistics for 2016, grouped by FIPS code.

Results

The following results show both the visual relationship between obesity, diabetes and the Adjusted Gross Income (AGI). The visual graph, showcasing data projected onto the United States map, was made using Plot.ly. Polt.ly is a data visualization software with packages available for Python. The package was utilized for generating the maps based on FIPS codes. Several graphs were generated for the different parameters we gathered.
Figure 1 and 2 show obesity percentages for the years 2008 and 2013, respectively:

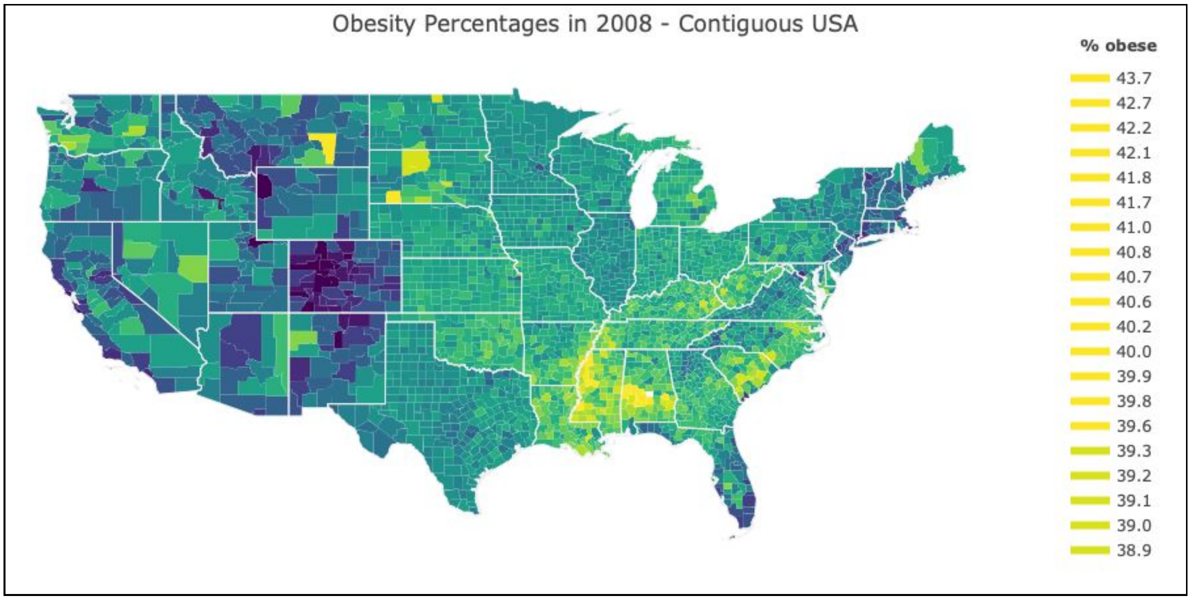

Figure 1: Obesity percentages for 2008

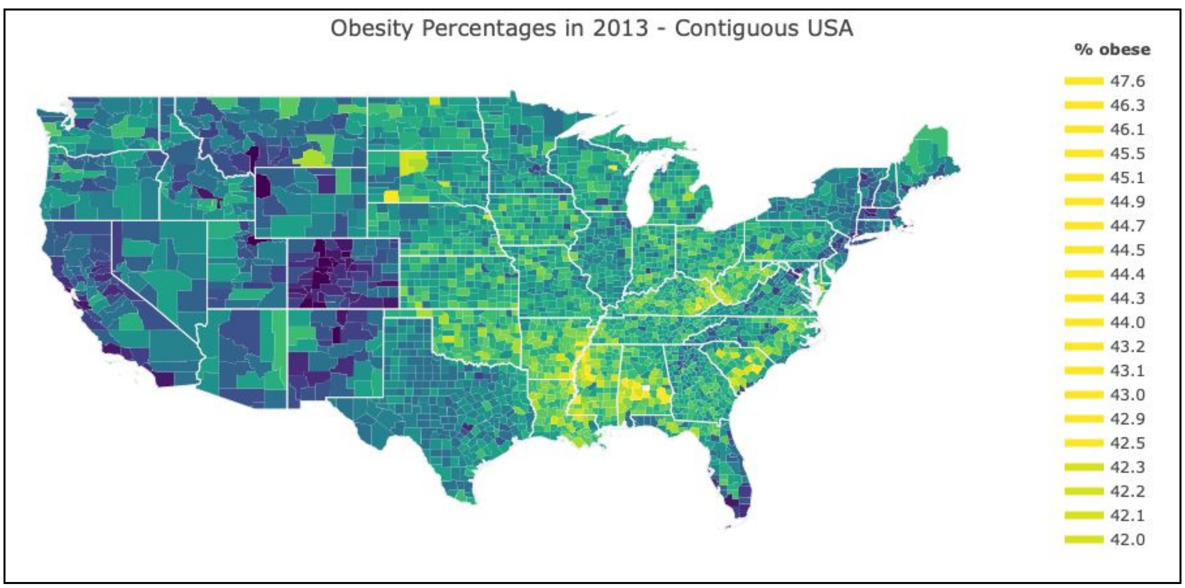

Figure 2: Obesity percentages for 2013

From figures 1 and 2, it is shown that obesity rates are the highest in southern states' counties. Alabama, Louisiana, and west Virginia show the highest obesity percentages in the contiguous United States. It is also shown that there has been a significant increase in obesity rates from 2008 to 2013. Further comparisons are made with the next figures. Figures 3 and 4 display diabetes percentages for the years 2008 and 2013, respectively:

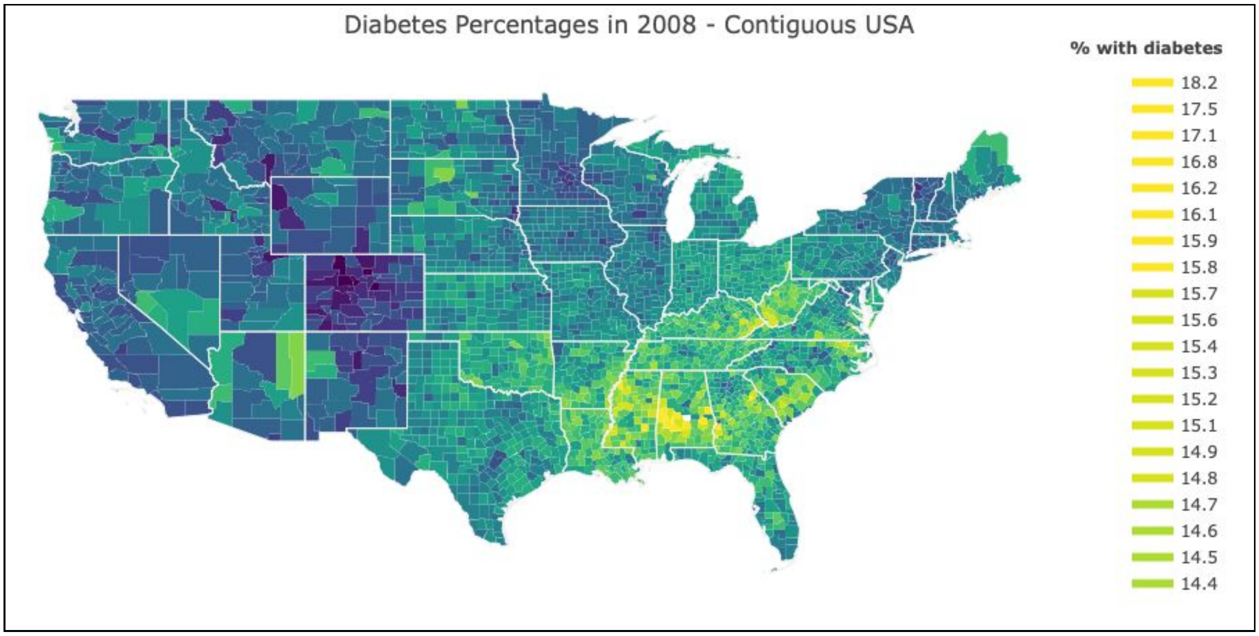

Figure 3: Diabetes percentages for 2008

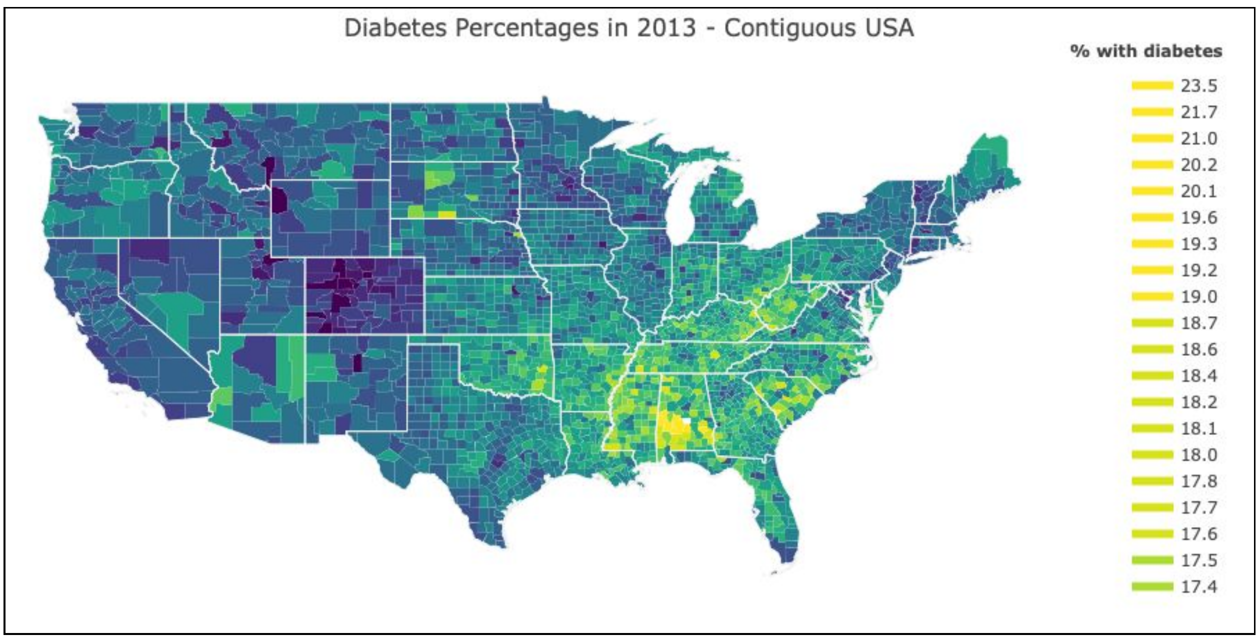

Figure 4: Diabetes percentages for 2013

Comparing the diabetes visualised maps with the obesity ones, it is shown that there is a positive correlation between the two health parameters. Obesity and diabetes percentages were the highest in the southern states, mainly Alabama and Louisiana. Figure 5 gives a numerical version of the above visual graphs. Figure 5 on the next pages shows the positive correlation between Obesity versus diabetes in 2013.

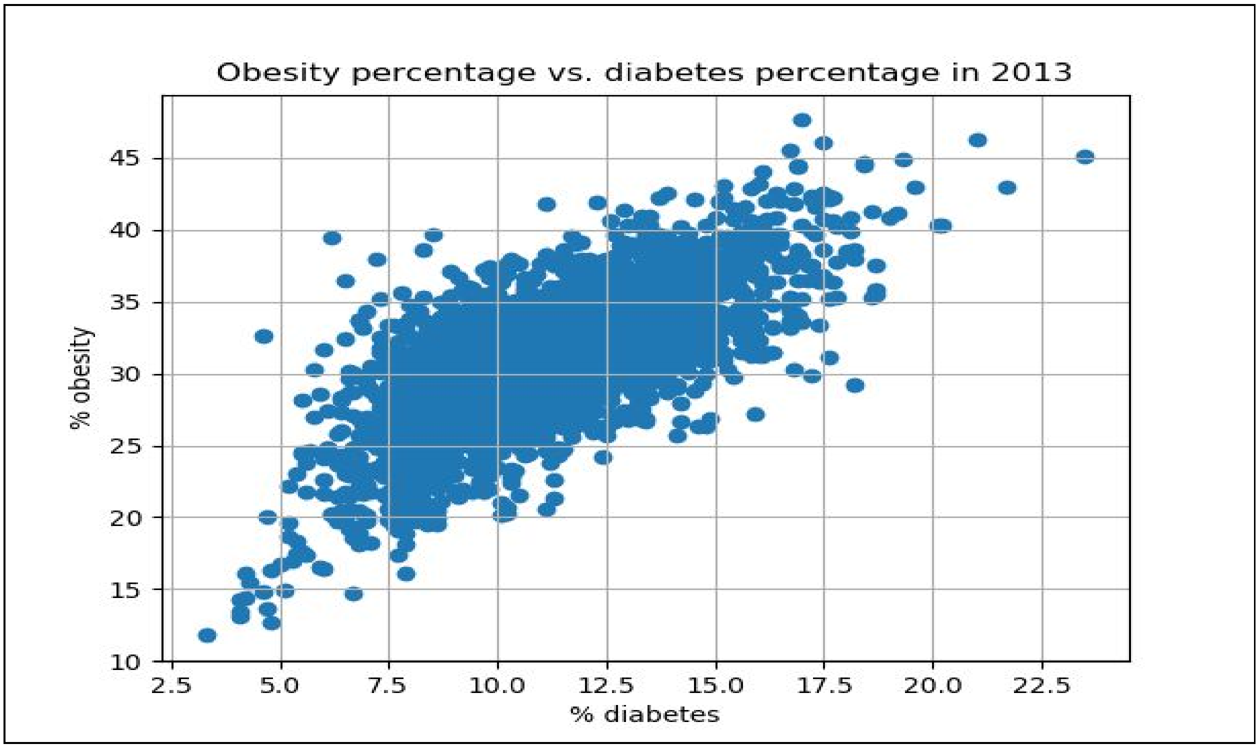

Figure 5: Obesity vs. diabetes percentages for 2013

Based on the previous 5 figures, we found that obesity and diabetes are positively correlated. States that had a high diabetes percentages also had a high obesity rate. This suggests that comparing a county's income level with any of these health parameters gives a general understanding of income vs. health evaluation. In other terms, Adjusted Gross Income (AGI) correlation with health parameters are expected to be similar. Figure 6 below showcases the Adjusted Gross Income (AGI) projected values on the United States' map:

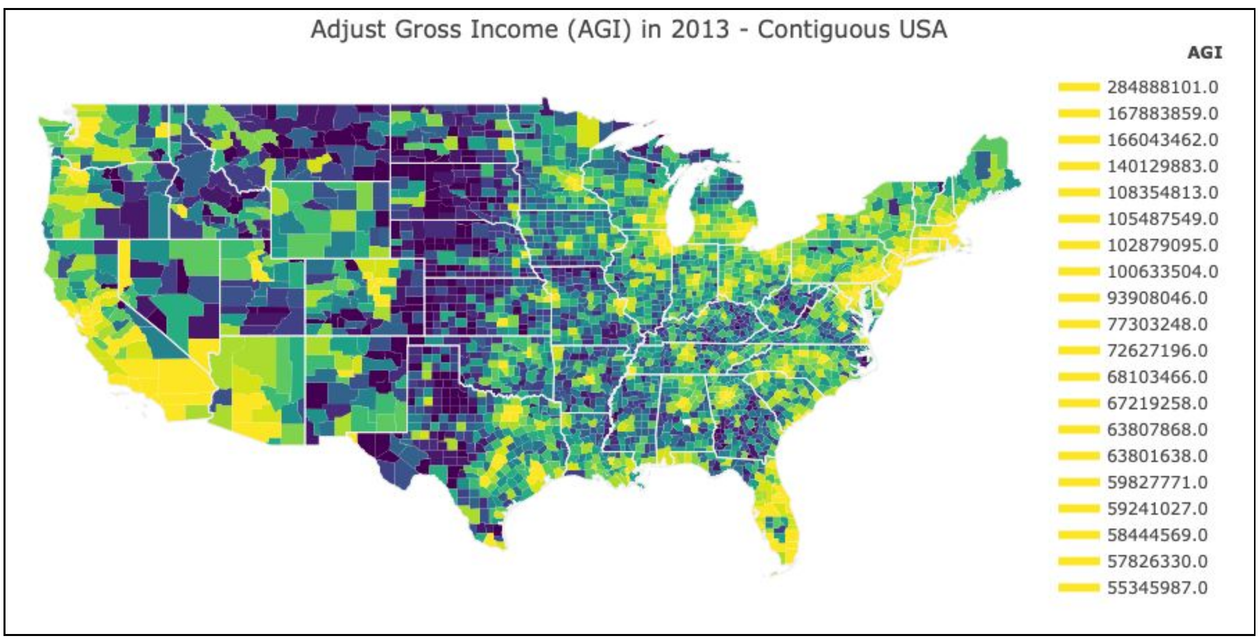

Figure 6: Adjusted Gross Income (AGI) values per county for 2013

Figure 6 shows the adjusted gross income distribution across the contiguous United States for 2013. Based on figures 2, 4, and 6, there is no apparent visual correlation between health parameters and adjusted income level per county. The counties with the highest adjusted income levels also house major cities in the United States. Taking the state of Texas as an example, counties where Houston, Austin, and Dallas cities are located have the highest adjusted income within Texas. Those cities also have a higher population than other counties in Texas. This suggests that adjusted income level is directly related to population levels. Further analysis is required to proof this correlation. Figures 7 and 8 show the relation between the Adjusted Gross Income (AGI) versus Obesity and Diabetes, respectively:

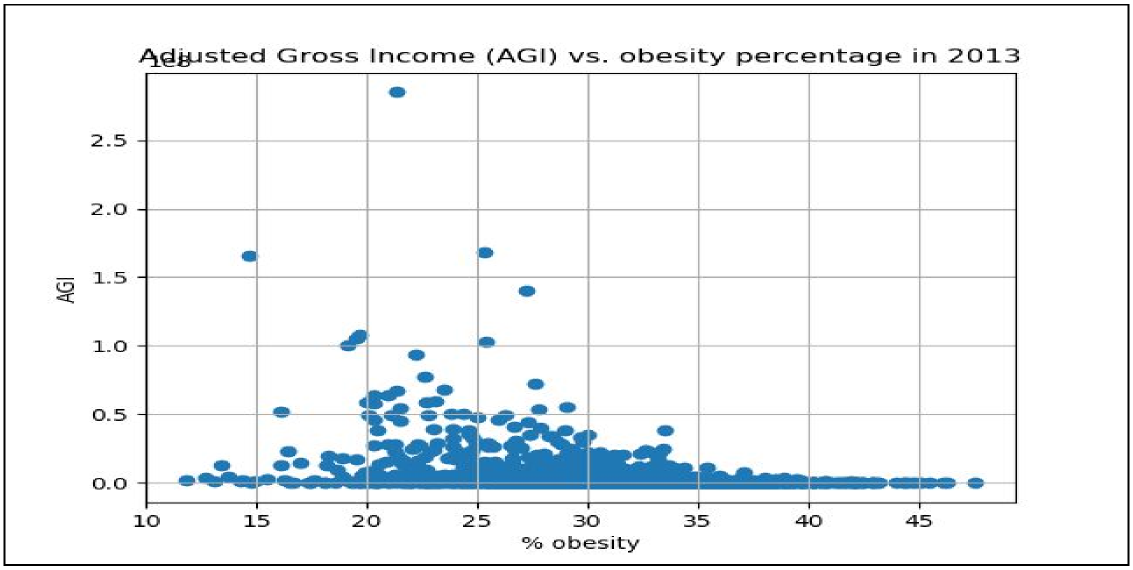

Figure 7: Adjusted Gross Income (AGI) versus obesity percentages for 2013

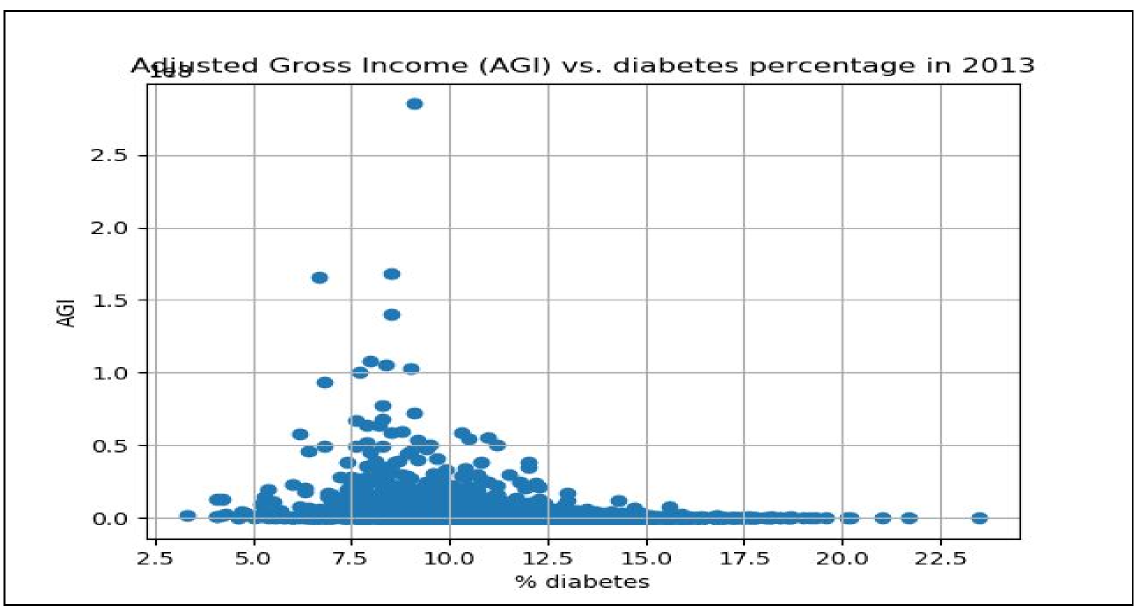

Figure 8: Adjusted Gross Income (AGI) versus diabetes percentages for 2013

While health parameters showed a positive correlation with each other, based on figures 7 and 8, county wise, there appears to be no noticeable correlation between adjusted income level and obesity or diabetes. Based on figures 7 and 8, the graphs vaguely show a normal distribution of health parameters with the adjusted income values. On average, counties at the 50th obesity percentile had higher adjusted income level. The same is true for diabetes values, were the 50th percentile also had measurably higher adjusted income levels.

The conclusion section offers more insight on these graphs, as well as recommended future areas of work

Conclusion

Based on the generated graphs, and analyzed datasets, no measurable correlation was found between the two health parameters (obesity, diabetes) and the Adjusted Income Level (AGI). The figures showed a normal distribution of obesity and diabetes values versus AGI. Low and high obesity rates appeared to have low income levels, while obesity rates at the 50th percentile had significantly higher adjusted income level. Diabetes rates on the two extremes also had the same behavior. Low and high diabetes rates had low adjusted income level, while the 50th percentile had higher adjusted income level.

One recommended area of improvement is to describe how minorities population percentages in the United States affect health levels in a county. Observation of race/minorities income levels and population percentages is also recommended. Whereas this report did not list race population percentages, we believe a deeper analysis is needed to determine whether race affects the health level or income level. Other factors to consider are: gender, immigration status, education level, family size, and age.

Another improvement area is to include more health and income parameters in the study. In this study, we only analyzed obesity, diabetes, and Adjusted Gross Income (AGI) data parameters, and we only described their correlation on a county level. Another recommended approach is to measure whether a underlying correlation exists on an individual level. As in, whether an individual's health level is affected by his\her's direct income level.

References

[1] Kawachi, I., & Kennedy, B. P. (1999). Income inequality and health: pathways and mechanisms. *Health service research*, 34(1 Pt 2), 215-27.

[2] Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences of the United States of America*, 107(38), 16489-93.

[3] Wagstaff, Adam & Doorslaer, Eddy. (2000). Income Inequality and Health: What Does the Literature Tell Us?. *Annual review of public health,* 543-67.

[4] Bryan, Kevin A. and Martinez, Leonardo, On the Evolution of Income Inequality in the United States (2008). *FRB Richmond Economic Quarterly*, Vol. 94, No. 2, Spring 2008, 97-120.

[5] L Shavers, Vickie. (2007). Measurement of Socioeconomic Status in Health Disparities Research. *Journal of the National Medical Association*. 99, 1013-23.

[6] Marmot, Michael. (2002). The Influence Of Income On Health: Views Of An Epidemiologist. *Health affairs (Project Hope)*. 21, 31-46.