

Relation Between Income and Health in the US

Team 8 Project 1 Report

Sayeed Alvi, Ashish Kumar, Mason Rumuly, Kanchan Satpute

Abstract

This report examines the relation between income and health through a study of food indicators and income data from past years. There are two major components of this study which are; bringing the relevant data together by conversion of ZIP code to FIPS code and analyzing the income correlation for further study. Our analysis supports the claim that individual income is positively correlated with health as well as uncovering some notable exceptions and interesting relationships with health care.

I. INTRODUCTION

CORRELATION of health to income level is relevant to many questions of political economy, and is often used as a political football to speak to issues of equity, of the effectiveness of modern medicine, and of life in a modern culture. When so many talking points rest on one empirical claim, and in a state of media where any and every statement is attacked by opposing pundits, it is important to check the validity of that claim. We investigate the claim that income level is positively correlated to health in the United States of America in order to support or refute it, avoiding implicating a causative link or suggesting policy in order to avoid biasing our results.

For our analysis, correlation will be defined as the Pearson linear correlation coefficient r . Because the sample size was immense even after removing invalid data or data for locations which were not included in all data sets, and a large fraction of the population, the single-tailed probabilities of incorrectly rejecting the null hypothesis were uniformly vanishingly small, on the order of 10^{-200} . Therefore, they will not be referenced in this analysis.

A. Food Atlas

The Food Environment Atlas is a set of food environment statistics compiled by the United States Department of Agriculture "to assemble statistics on food environment indicators to stimulate research on the determinants of food choices and diet quality, and to provide a spatial overview of a community's ability to access healthy food and its success in doing so" [1]. The statistics are provided at a county level indicated using Federal Information Processing Standards (FIPS) county codes, which provides a reasonably fine granularity of health indicators without breaking individual privacy. It includes results of studies from multiple years, notably median income data from 2015 as well as adult diabetes and obesity rates for both 2008 and 2013. It is not obvious that the geographic distribution of income would be consistent enough over time to make any results valid. This question of application is addressed using income data from the Internal Revenue Service (IRS).

B. IRS Income Data

The IRS provides statistics related to the annual collection of the Federal Income Tax with a Zip Code geographic granularity [2]. It includes the Adjusted Gross Income (AGI) total for each Zip code. Since a count of returns, dependents, and exemptions are included, this data also provides a reliable population estimate which we use to weight analyses of correlation in the data; it also enables us to use the AGI to calculate the mean income for each given Zip Code. The drawback of using this mean directly is that it is less informative about individual status than the median provided by the Food Atlas, dominated as it is by large-income outliers on a single-tailed distribution. Therefore, we use this data to qualify the use of the 2015 median income data with health information from other years. This information is aggregated into FIPS counties in order to use with the Food Atlas data.

C. FIPS and ZIP

As the data has been collected from multiple sources, aggregation of the data is a significant task so as to produce meaningful data for our analysis. The ZIP-county-FIPS data-set [3] provides us with the look-up between US ZIP code, FIPS State and County IDs to County/State. We use this data to map the ZIP code from the income data to the FIPS code from the food data. Essentially, for each FIPS code in the food data, all the ZIP codes were extracted from the mapping data. The income statistics corresponding to these ZIP codes were then aggregated to represent according to each FIPS code. Any ZIP code which belonged to multiple counties was included in the estimation of the average income for all related ZIP codes.

D. County Health Report Data

Other potential data sources were identified to establish the relationship between income level and health. The University of Wisconsin Population Institute and Robert Wood Johnson Foundation collaborated to start The County Health Rankings & Roadmaps program which helps communities identify and implement solutions that make it easier for people to be healthy in their neighborhoods, schools, and workplaces. It collects data on health behaviors, socio-economic factors, clinical care and physical environment from different government organizations and ranks all counties based on weighted parameters. In addition to this additional Community Health Status Indicators (CHSI) like infant mortality rate, child mortality rate, smokers, food Insecurity etc. to combat obesity, heart disease, and cancer were also included in the study.

II. GEOGRAPHIC INCOME

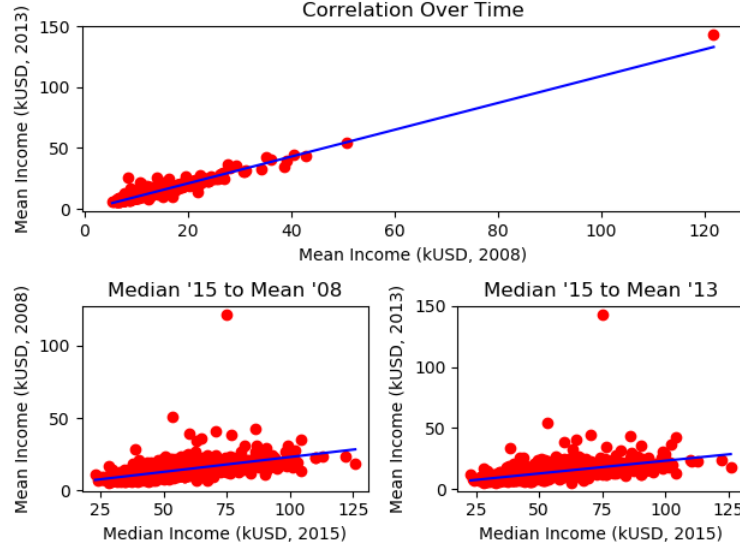


Fig. 1: Correlation Of Income

In order to ensure the applicability of the median income information from 2015 provided by the Food Atlas could be applied to the 2008 and 2013 health indicators, we looked at how much the geographic distribution of income changed over time. Figure 1 shows our findings; mean income in 2008 is very strongly correlated with mean income for 2013 when weighted with average population between the two years at $r = 0.9896$. Also included in our analysis is a sanity check on the relevant mean and median income; the two are expected to correlate in a less strong fashion for the single-tailed distribution involved, since an underlying identical distribution of income would not produce 1 as the maximum possible value of $|r|$. With that in mind, the strong correlation observed with $r = 0.572$ for 2008 and $r = 0.538$ for 2013 strongly indicates that the 2015 median income data generalizes well over time with a manageable amount of noise.

This result accomplish two goals. First, it allows for claims to be made about contemporaneous health and income data even if the exact years available for each dataset are not identical. Second, the lack of variance in geographic income distribution over time implies that they are permanent enough to be or produce consistent effects if they exist.

III. RELATION TO HEALTH ACCESS MEASURES

Household Income is responsible for shaping our lifestyle choices. We identified the relation between median household income and ease of access to certain factors like access to exercising opportunities, access to healthy foods and the Access to Dental care in Fig. 2. We can see that wealthier people have greater access to places for physical activities like Fitness Studios, etc. Food environment index is an indicator of access to healthy foods -0 is the worst and 10 is the best. From the figure, wealthier people seem to have higher access to healthy foods. The figure also shows the percentage of population that is food insecure i.e lacking reliable access to a sufficient quantity of affordable, nutritious food. We can analyze that wealthier people have lower food insecurity than the less wealthier ones. All these factors play a major role in the health of a person and ease of access to these resources will be necessary to improve their health.

Health care provider and institutions play a important role in upkeep of health of the general public. As per data set, all counties have fair ratio of population to medical professionals. It is evident from the Fig. 3 that there is a weak relation between the health care providers and income, thereby each county has access to health care providers irrespective of income level.

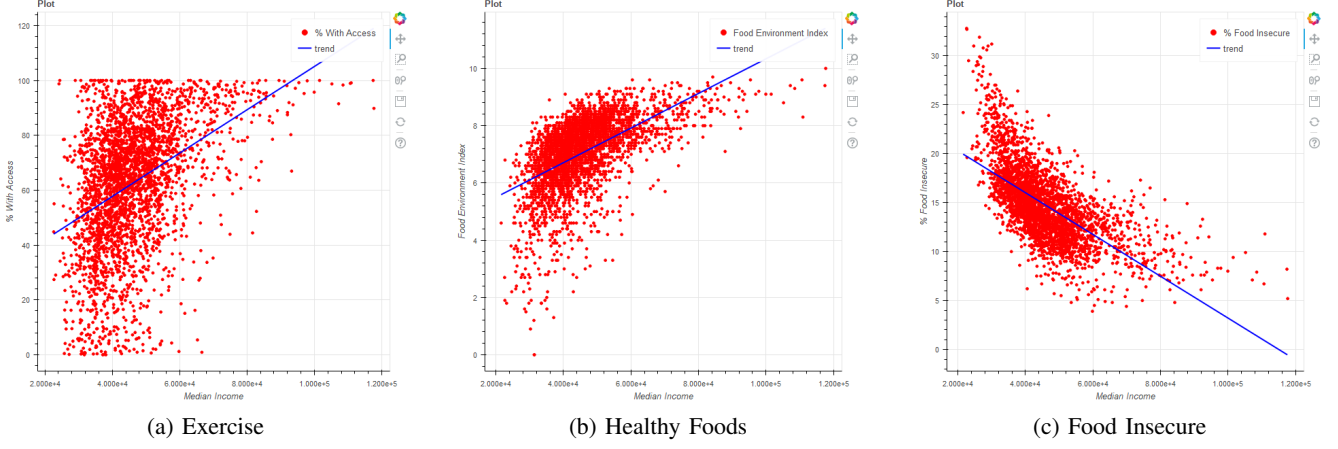


Fig. 2: Income vs Access to Exercise Opportunities, Access to Healthy Foods and Percentage of Food Insecure Population

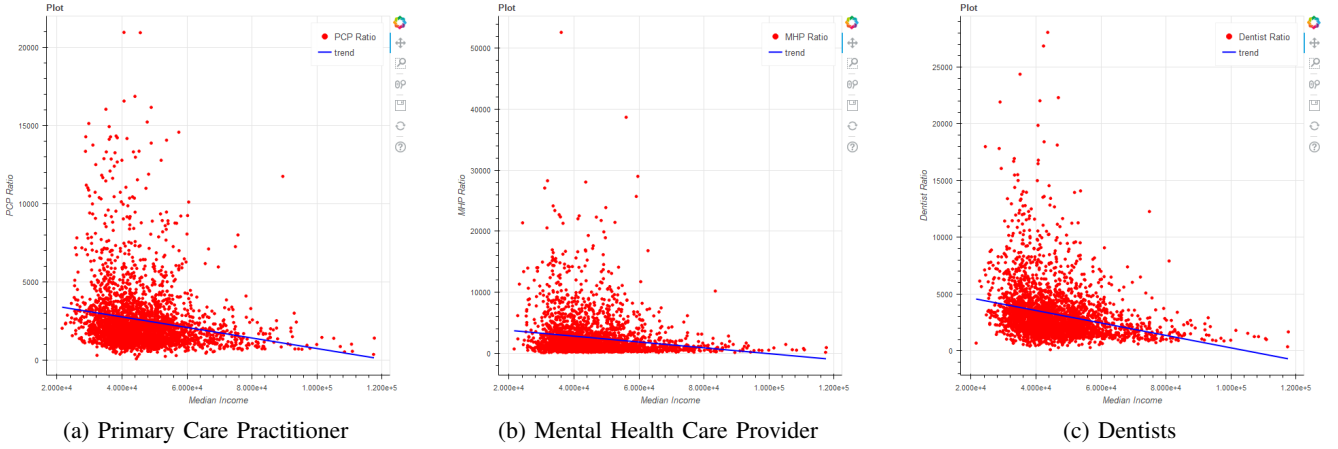


Fig. 3: Income vs Primary Health Care, Mental Health Practitioner, Dentists Ratio

IV. RELATION TO COMMUNICABLE AND NON-COMMUNICABLE DISEASES

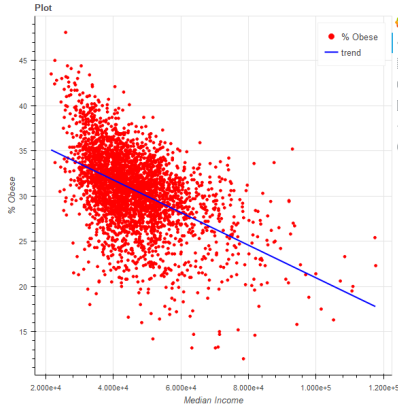
Health includes physical, mental and social well being. Physical well being can be affected by incidence communicable or/and non-communicable diseases. Mental health can be affected by presence of mental disease, depression etc. The World Bank recognizes other factors which are related to health i.e. Infant mortality rate, child mortality rate etc. Some of these health indicators have direct related to health while others are indirect. Data sets [4] provide useful information on various health indicators of counties in USA. Here we will analyze each of these indicators to see if they are related to income.

It is observed that all counties have high rates of obese and diabetic person. Obesity and diabetic leads to several non communicable diseases like cardiovascular diseases. It can be observed from the Fig. 4a and Fig. 4b that low median income counties have higher obesity and diabetic population. There is clear trend of poor health with lower income.

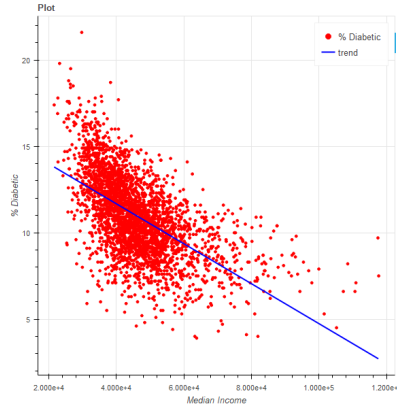
Cigarette smoking harms nearly every organ of the body, causes many diseases, and reduces the health of smokers in general [5] [6]. There is a trend of more number of smokers with lower income thus counties with more smokers will likely to have poor health from Fig. 4c .

Childbearing by very young mothers is a matter of public concern because of the elevated health risks for these mothers and their infants [7]. As per Centres of Disease Control and prevention infants with low birth weight i.e. less than 2.5 lbs at birth are known to have health risks. Teen birth rate and low birth rate is somewhat more linked to infant mortality. These all are indicators of poor health. As per Fig. 5a, 5b and 5c trend shows there is a correlation between income and IMR, low birth weight and teenage mothers.

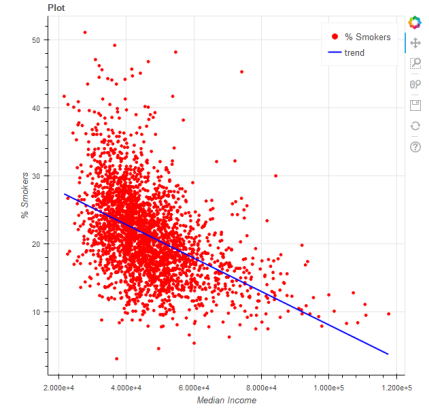
Though most of the health indicators discussed previously are strongly related to income, some indicators are weakly or not related. HIV and Chlamydia, which are sexually transmitted diseases, do not appear to be strongly related to income as observed in Fig. 6a and 6b. Another issue of excessive alcohol consumption which may lead to liver, heart issues is seen to higher among counties with more income as seen in Fig. 6c.



(a) Obesity

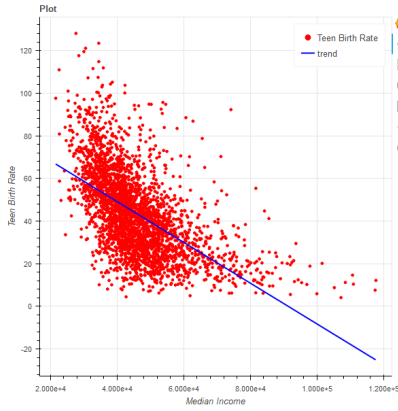


(b) Diabetic

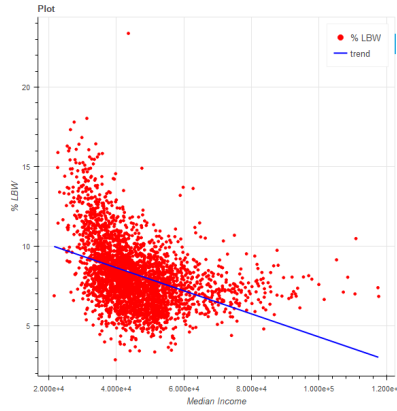


(c) Smokers

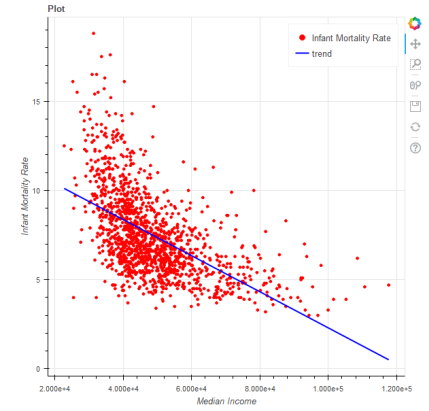
Fig. 4: Income vs Obesity, Diabetic, Smokers



(a) Teen Birth Rate

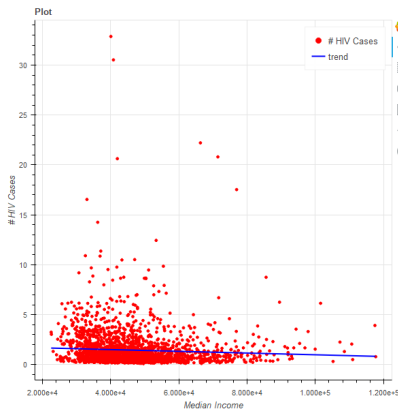


(b) Low Birth Weight

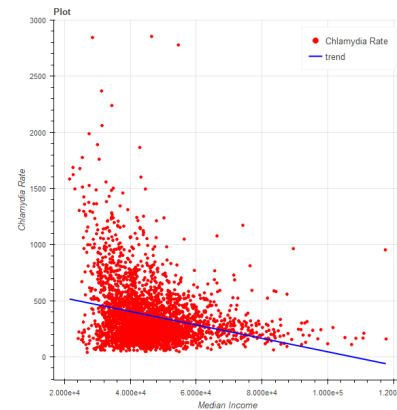


(c) Infant Mortality Rate

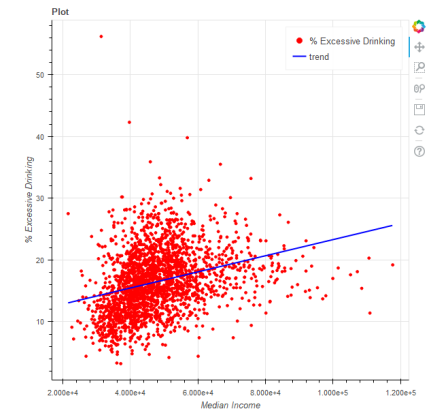
Fig. 5: Income vs Teen birth, Low Birth Weight, Infant Mortality Rate



(a) HIV cases



(b) Chlamydia Rate



(c) Excessive alcohol drinking

Fig. 6: Income vs STDs and Excessive drinking

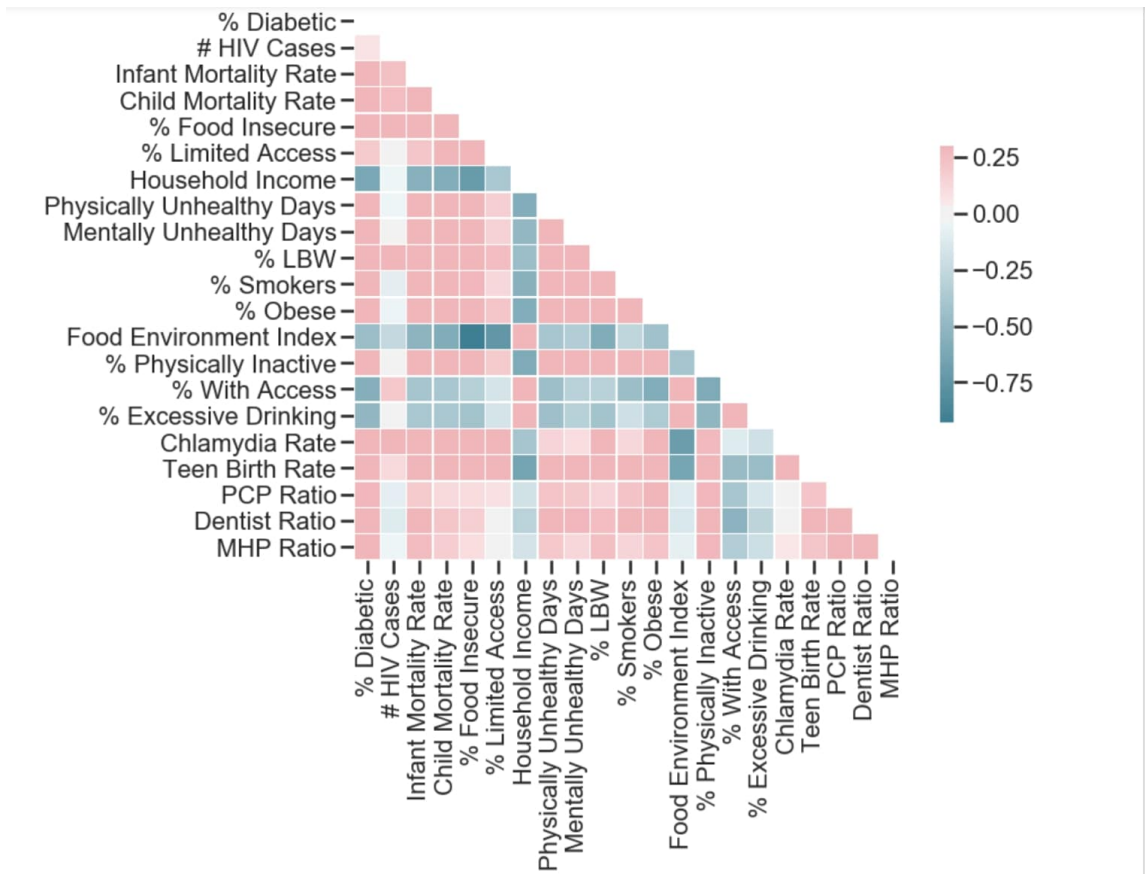


Fig. 7: Diagonal Correlation Matrix

V. STATISTICAL ANALYSIS OF HEALTH INDICATORS VS INCOME

Diabetes, Obesity, arthritis, coronary heart diseases and many other health problems affect the population in each income bracket. After studying the correlation between some of the health problems with income, we have observed that certain problems are less widespread among wealthier people. The correlation matrix in Fig. 7 gives us an idea of the dependency of various health factors on household income. The blue color shows that the correlation is highest and as the color fades to pink, the correlation decreases. For Household Income we can see that the percentage of diabetic or obese people, physical & mental unhealthy days, number of HIV cases are more correlated. This validates our take on how health problems affect more people from low income bracket.

VI. CONCLUSION

Our analysis of the income and health data for the US supports the claim that income is positively correlated with health and access to food and participation in exercise in most categories, though there are notable exceptions: HIV cases are not significantly correlated with income, and excessive alcohol consumption increases with income. However, the number of health-care providers per population is negatively correlated with the income level, and negatively correlated with health. These interactions point to opportunities for very interesting future investigation scope to find the causal relationships involved.

REFERENCES

- [1] *Food Environment Atlas*. United States Department of Agriculture. 27 March 2017. <https://www.ers.usda.gov/data-products/food-environment-atlas/>.
- [2] *Individual Income Tax Statistics - ZIP Code Data*. Internal Revenue Service. 30 Aug 2018. <https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi> <https://www.ers.usda.gov/data-products/food-environment-atlas/>.
- [3] <https://www.kaggle.com/danofer/zipcodes-county-fips-crosswalk/home>
- [4] <http://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation/national-data-documentation-2010-2016>
- [5] U.S. Department of Health and Human Services. The Health Consequences of Smoking 50 Years of Progress: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014 [accessed 2017 Apr 20].
- [6] U.S. Department of Health and Human Services. How Tobacco Smoke Causes Disease: What It Means to You. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2010 [accessed 2017 Apr 20].
- [7] Menacker F, Martin JA, MacDorman MF, Ventura SJ. Births to 1014 year-old mothers, 1990-2002: Trends and health outcomes. National Vital Statistics Reports; vol 53 no 7. Hyattsville, Maryland: National Center for Health Statistics. 2004.