DOES WEALTHIER MEAN HEALTHIER?

Divyank Garg, Samarth Mistry, Ranjith Tamil Selvan, Priyadharsheni Balasubramanian

Health is a state of complete wellbeing in all aspects of an individual's life - physical, social and mental. It is not simply the physical wellbeing or the absence of disease. There are several factors that can affect a person's health and they can be broadly categorized under food habits, environmental factors, social conditions, economic conditions and interactions with external factors. A few notable ones that came up first during the authors' discussion were food availability, work conditions, commute conditions, stress and healthcare availability. A change in any of the above-mentioned factors can impact the factors. Every factor is interlinked and interdependent on one another.

According to the monthly labor review article release in February 2017, the life expectancy of an American has increased over the last 50 years. But the average "goodness" of health among the population has fallen dramatically and the self-reported "health-scores" by individuals as either "good" or "very good" has also decreased. This has been coupled with increasing rates of obesity and diabetes among the general population.

The need of the hour is to look at actual numbers backed up by competent research and analysis. This will be helpful in framing policies that could promote a better lifestyle; thus, increasing the chances of bringing up the health quality index of the population.

Through this article, we try to shed some light on the relationship between the wealth and health of the people. While some of you might deny the benefits of large amount wealth, it has several effects on the physical and the mental wellbeing of the individual. Apart from the direct advantages like having access to better quality food resources, better healthcare and a quality lifestyle, the effect of income on other psychological factors also can contribute massively to the health state of an individual.

Let's talk data

Our first wealth parameter is 'income'. We obtained this data from the publicly available data from the IRS for 2016. These are grouped by ZIP code and we extracted the income number as a per capita income for a county.

The health parameters have been obtained from the Food Atlas data that is provided by the U.S. Department of Agriculture. The statistics provided contains food environment indicators and provides spatial overview of access to healthy food. The major data points are obesity and

diabetes rates which have been used as direct indicators health. The Foot Atlas data is spatially grouped with FIPS code across United States.

Federal Information Processing Standard (FIPS) is a publicly announced standard that is developed by United States federal government for use in computer systems. This code uniquely identifies the counties and county equivalents in the United States and other territories of United States. The two datasets - the IRS dataset and the Foot Atlas datasets are combined over the FIPS code. The ZIP code data in the IRS data was converted to the equivalent FIPS code by using the county and ZIP code information and a mapping of this information to the FIPS code.

The combined dataset was then used to perform a statistical analysis to inspect the relationship between income and health indicators.

Initial Observations

Now, we have got important variables like per capita income, diabetes & obesity and would find the variation of its values based on county wise in the US map.

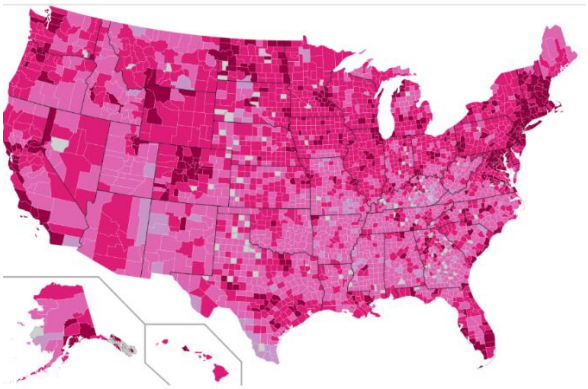


Fig: Average per capita Income value variation based on US county

Here the light pink color tells about the county with less per capita income value and the dark pink color tells about the county with high per capita income value. Different shades of pink from light to dark tells about the variation of income from low to high.

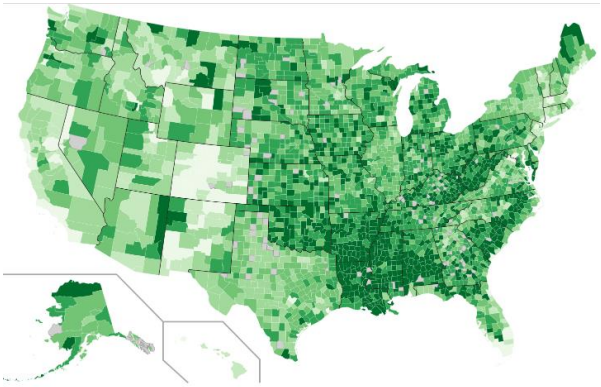


Fig: Obesity rate variation based on US county

Here the light green color tells about the county with less obesity rate and the dark green color tells about the county with high obesity rate. Different shades of green from light to dark tells about the variation of obesity rate from low to high.

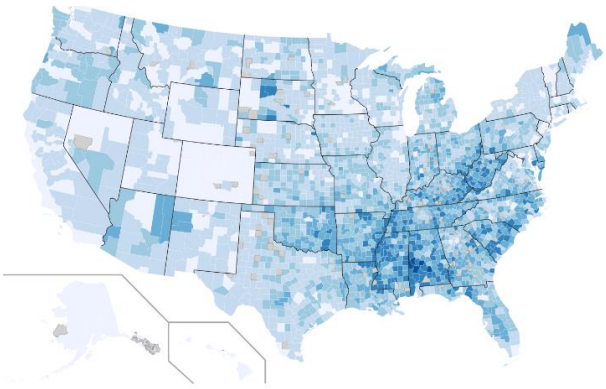


Fig: Diabetes rate variation based on US county

Here the light blue color tells about the county with less diabetes rate and the dark blue color tells about the county with high diabetes rate. Different shades of blue from light to dark tells about the variation of obesity rate from low to high.

We can observe from the above figures that higher income was related with lower diabetes and lower obesity rates. The per capita income of New York county and Teton county for the fiscal year 2016 was about $130000 and $119000 respectively, which is the highest reported per capita income, according to stats on the IRS Website (SOI Tax Stats). New York County was around 14.7% and 6.7% respectively, and for Teton County, 12.7% and 4.8% respectively, according to the stats found on the Food Atlas. The per capita income of Wadehampton county and Clay county for the fiscal year 2016 is found to have the lowest reported per capita income, around $10000 and $11000 respectively. The corresponding percentages of people with obesity and diabetes for Wadehampton county was around 38% and 14.1% respectively, and for Clay county, 32.6% and 4.6%. Thus, the difference in obesity and diabetes percentages between county with highest per capita income (New York) and county with lowest per capita income (Wade Hampton) is roughly around 23.3% and 7.4% respectively.

We could make the same analogy with the counties with highest and lowest diabetes percentages. It is observed that counties Lowndes and Perry have the highest diabetes percentages of 23.5% and 21.7%. The corresponding per capita income of these counties have low figures of $18000 and $19000 respectively. Counties Eagle and Routt exhibit the lowest diabetes percentages of 3.3% and 4.1%, with corresponding per capita income of $57000 and $58000 respectively. Obesity percentages show similar patterns too. Counties Claiborne and Greene have the highest obesity percentages of 47.6% and 46.3% respectively. The per-capita incomes of these counties were observed to be $17000 and $16000 respectively. The counties Eagle and Teton, having the lowest obesity percentages of 11.8% and 12.7% have per capita income of $57000 and $118000 respectively. It would be interesting to find out whether there is a threshold above which additional income may no longer be associated with better health. Alternatively, there may or may not be a threshold below which further reductions in income do not harm health. The analysis also showed counties of states with higher per capita income – California and New York are associated with lower obesity and diabetes rates.

Finding Correlation between Important Variables

Here it can be seen that obesity and diabetes is correlated to income and that too inversely related. Apart from the income, there is also another variable related to obesity and diabetes like- SNAP, so its effect is also taken into consideration.

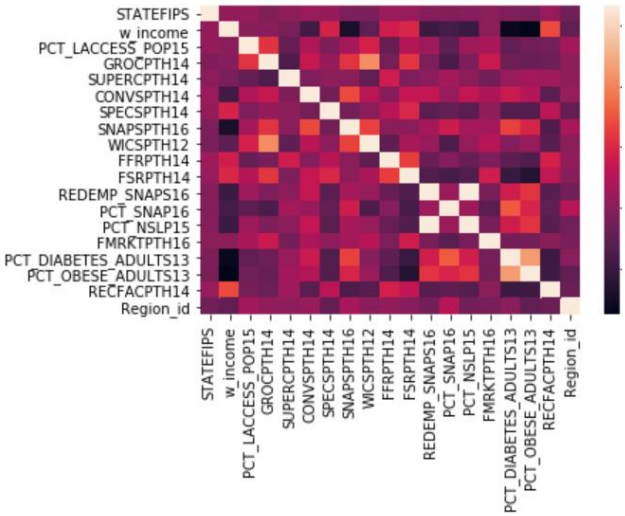


Fig: Heat map between variables for finding correlation

Finding relation between Income, Obesity and Diabetes across the US

The relation between obesity and per capita income can be visualized by making a scatter plot between them. These data points are for individual county throughout the US

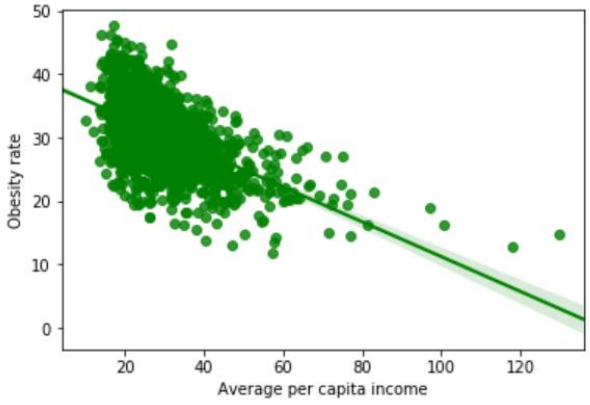


Fig: Relation b/w obesity and income

From scatter plot, it can be seen that as per capita income increases then the obesity rate decreases. It can be inferred from this that the people which have high income are less prone to obesity. The people with higher income are able to invest on things which are healthier even though they are costly. The correlation value between average per capita income and obesity was found to be -0.5298, which verify the negative and high relation exist between these factors.

Now looking this relation, the diabetes can also be checked with respect to per capita income. The same concept of scatter plot is used to find any relation between them. Here also the data points are for individually county throughout the US

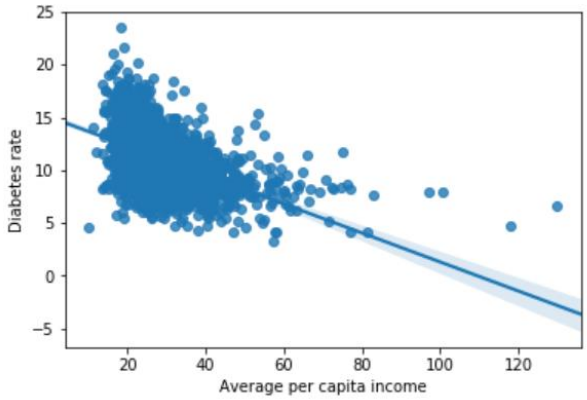


Fig: Relation b/w diabetes and income

Here it can be inferred that as the average per capita income increases for all county throughout US then Diabetes rate decreases.

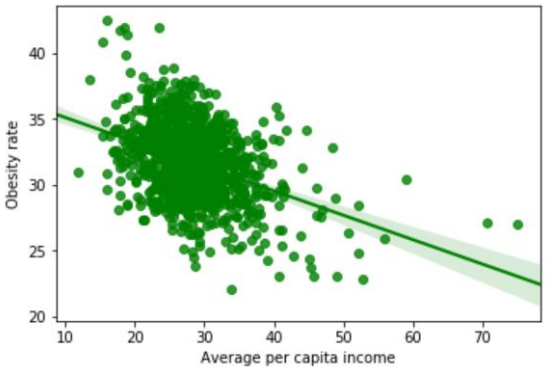
The correlation value between average per capita income and obesity was found to be -0.4830, which verify the negative and high relation exist between these factors.

Finding relation between Income, Obesity and Diabetes in US region as whole

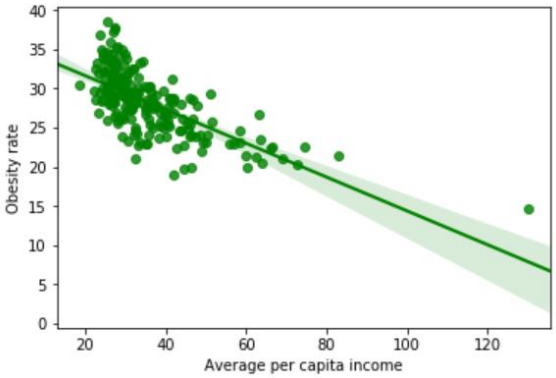
Now the individual county was grouped together region wise within US. As there are 4 major regions in US -Northeast,

Midwest, South, West, which comprise of state within them. The relation between obesity and average per capita income for region was found so that to know which region has strong relation between these two factor.

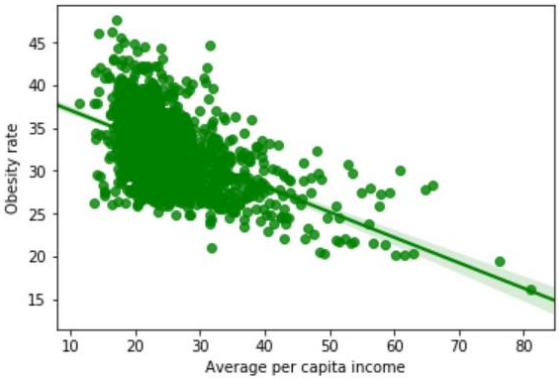
Fig: Relation b/w obesity and income for different regions in US



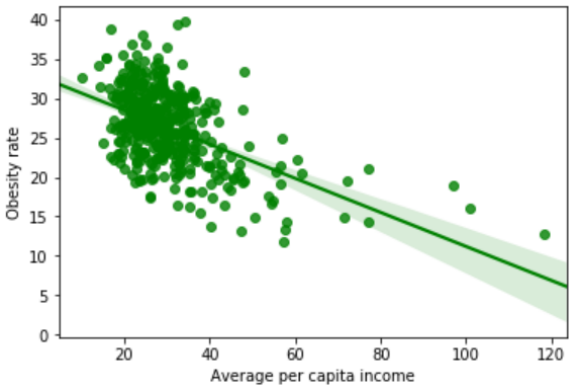
A) Northeast Region



B) Midwest Region



C) South Region



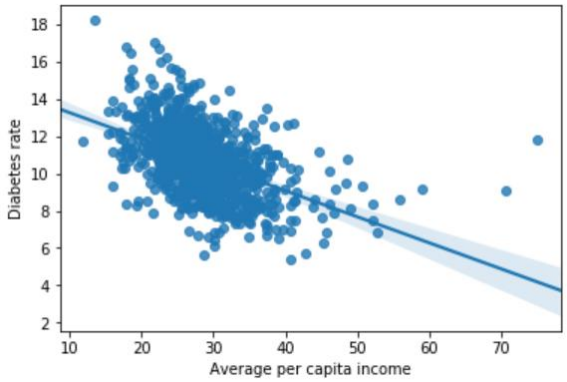
D) West Region

Region	Correlation between Obesity and per capita income
Northeast	-0.685
Midwest	-0.393
South	-0.524
West	-0.511

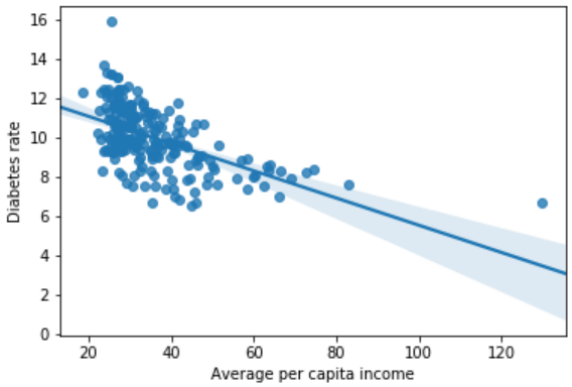
From table and graph, it can been seen that all are negative correlated which means in every region the more the income level is the less is the obesity rate. But in Northeast region it can be found that correlation value is highest among other region, so it means that the income plays a major role in deciding the obesity rate in that region.

Now, the same thing is done to find the correlation between diabetes and Avg per capita income for different region within US

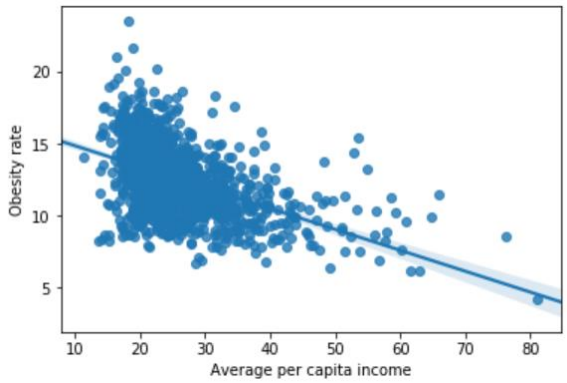
Fig: Relation b/w diabetes and income for different regions in US



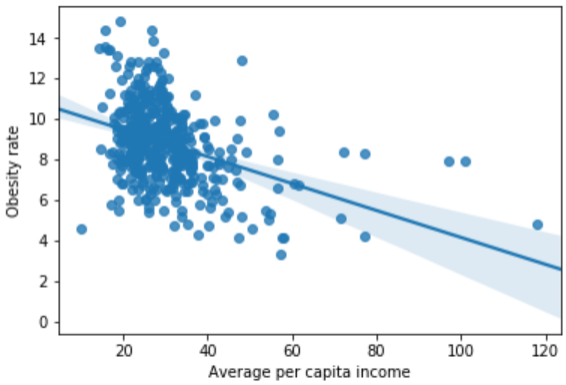
A) Northeast Region



B) Midwest Region



C) South Region



D) West Region

Region	Correlation between Diabetes and per capita income
Northeast	-0.559
Midwest	-0.472
South	-0.474
West	-0.385

From the table and graph, it can be seen that all are negative correlated which means in every region the more the income of people the more there is chance of getting diabetic. In Northeast region it can be found that the diabetes is highly correlated to income compared to other region which means income play important role in deciding diabetes rate in that region.

Finding additional variables to indicate wealth

Till now only income is taken important variable for knowing the wealth condition per county. There might be other variables which also contribute to the wealth factor for individual county so to find those other variable, random forest was performed.

In first case obesity was taken as target variable and based on it important variables were found based on importance percentage.

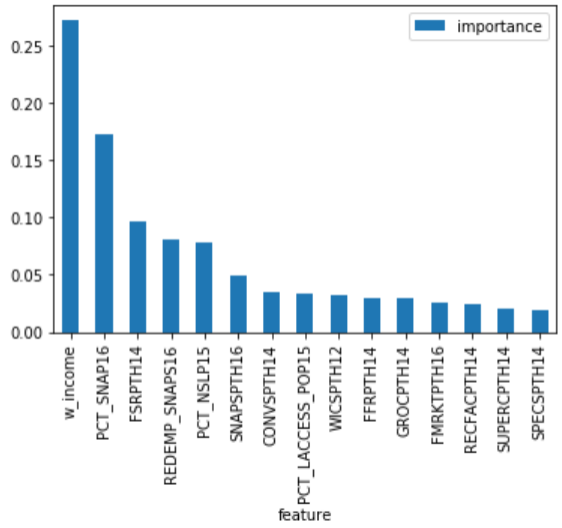


Fig: Important features related to obesity

Here income is the important factor related to obesity but PCT_SNAP16 is also one important factor after that. This graph is based on county wise for whole US.

In second case diabetes was taken into as target variable and based on it important variable were found based on importance.

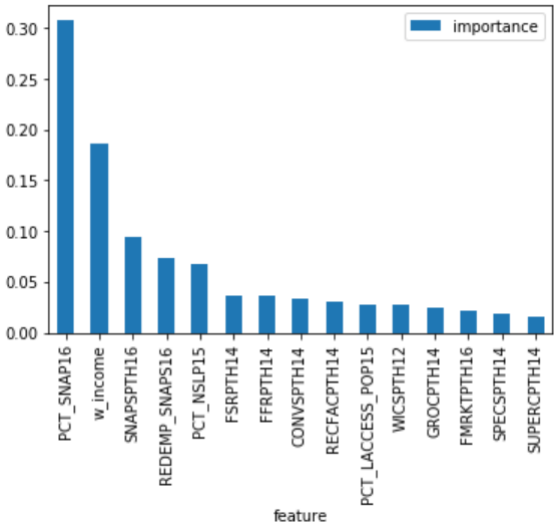


Fig: Important features related to diabetes

Here PCT_SNAP16 is the important factor related to obesity. This graph is based on county wise for whole US

so now based on this result the relation between PCT_SNAP16, obesity and diabetes can be easily found by scatter plot.

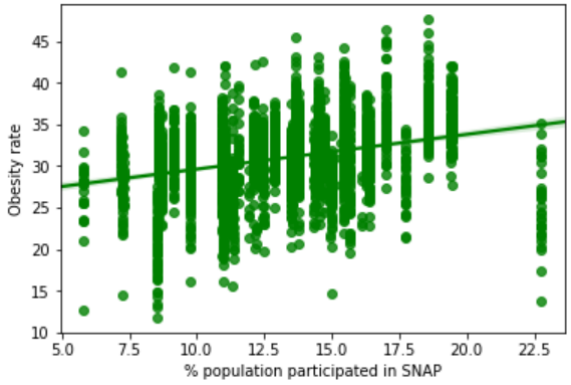


Fig: Relation between SNAP and obesity rate across the US

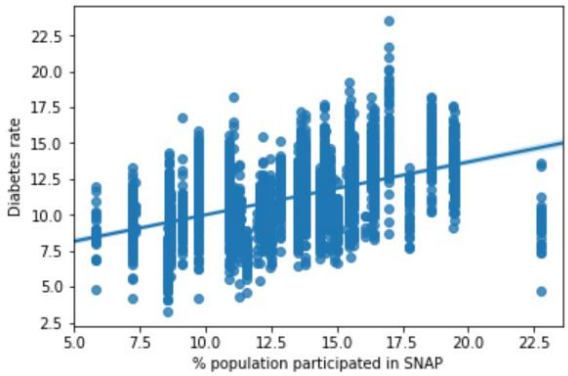


Fig: Relation between SNAP and obesity rate across the US

Here, as percentage of SNAP variable increases then the obesity rate and diabetes also increase. The correlation between these two variables can be found around 0.294 and 0.471 respectively. The SNAP variable tells about the percentage of people participated in the food assistance program so this means that the people those comes under this category

have low incomes. Hence, we find the increase in trend in above figures.

Conclusion

We can say that wealth and health are correlated to a good degree. Although we cannot conclusively say that this correlation also corresponds to causation. The causal relation can also be the other way around.

From this analysis, we would humbly submit that we would require additional data that characterizes wealth instead of just income. The reason is that very-high-income people can skew the per capita income of a county. Adding a new income like 'SNAP recipients' can give us richer information on wealth.

References

- https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/income-health-poverty-full.pdf
- https://www.bls.gov/opub/mlr/2017/beyond-bls/pdf/income-and-health-outcomes.pdf
- https://www.urban.org/sites/default/files/publication/49116/2000178-How-are-Income-and-Wealth-Linked-to-Health-and-Longevity.pdf
- http://www.health.state.mn.us/divs/opa/2014incomeandhealth.pdf