

# Capstone Project

**Premature Mortality as an Indicator to Measure the Health of People in US Counties**

**By Nalini Suresh**

## Why Does Health Matter?

*World Health Organization appropriately defines health as “Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.”*

---

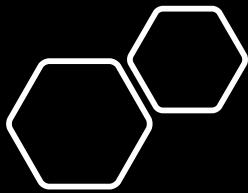
# How Healthy is Your Community?

---

## **How do we measure the health of the people in United States?**

Scientists link chronic diseases and other external factors that exacerbates health conditions such as diabetes, physical inactivity, access to mental health providers and people living with HIV. Other drivers include age factor, divide in living conditions between rural areas and urban areas of the country and gender contribute to premature death of a person.

Also, factors like median household income that is critical for families to be able to eat nutritious food, afford health insurance, house and education.



## What is Years of Potential Life Lost (YPLL)?

Age-adjusted Years of Potential Life Lost (YPLL)  
Before Age 75 per 100,000  
in all counties in United States

“Years of potential life lost (YPLL) is one measure of the impact of premature mortality on a population.”<sup>1</sup>

It measures the total number of years a person lost due to death below the age of 75 per 100,000 persons. Measuring premature mortality, rather than overall mortality, reflects the County Health Rankings’ intent to focus attention on deaths that could have been prevented. This is because some of the deaths can be avoided by making healthier life choice and strong public policies.

For example, using YPLL-75, a death at age 55 adds 15 to the sum while 78 adds 0 to the sum.

### *Method for calculating YPLL from a line listing*

**Step 1.** Decide on end point (65 years, average life expectancy, or other).

**Step 2.** Exclude records of all persons who died at or after the end point.

**Step 3.** For each person who died before the end point, calculate that person’s YPLL by subtracting the age at death from the end point.

$$\text{YPLL}_{\text{individual}} = \text{end point} - \text{age at death}$$

**Step 3.** Sum the individual YPLLS.

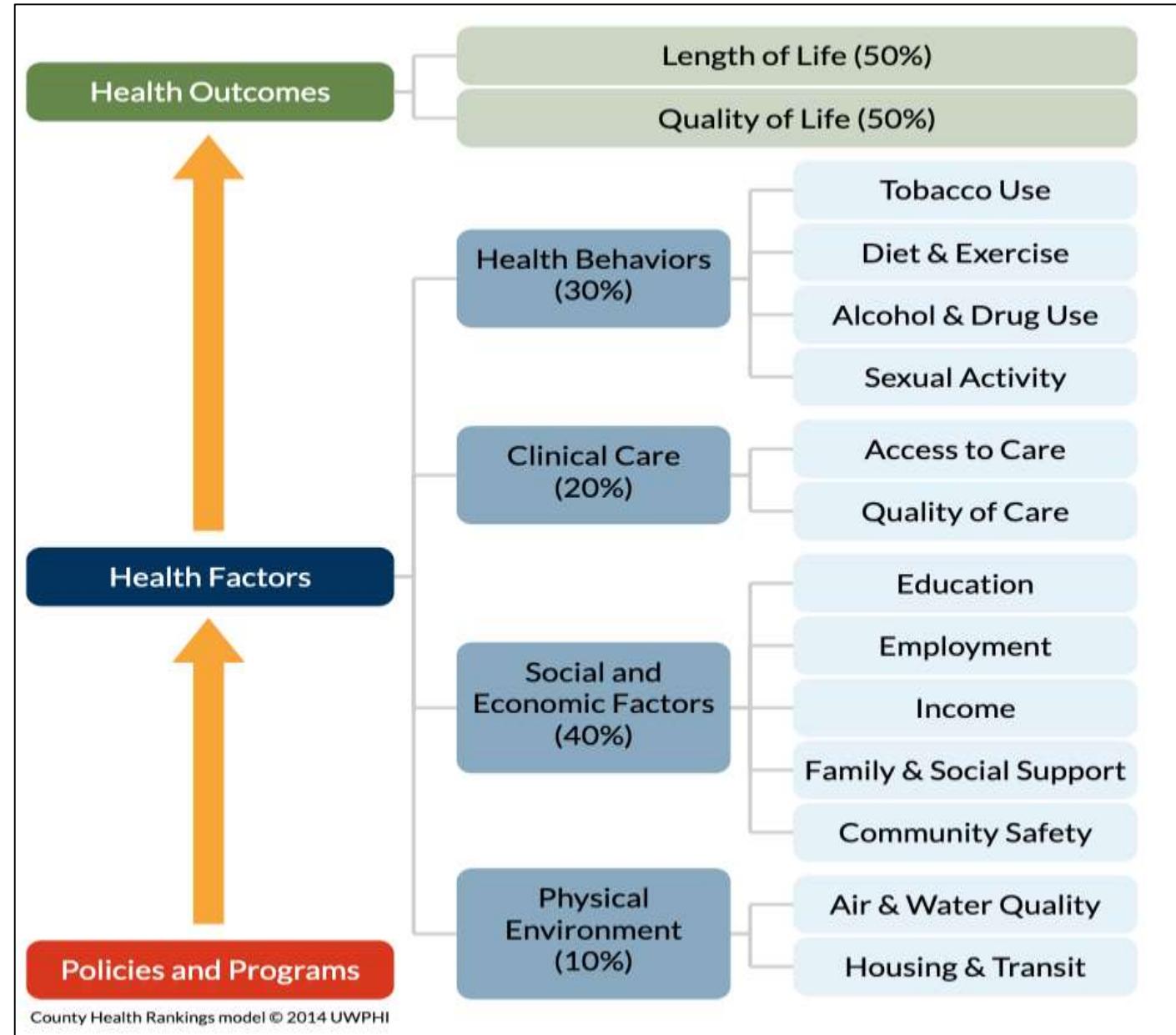
$$\text{YPLL} = \sum \text{YPLL}_{\text{individual}}$$

<sup>1</sup> <https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section3.html>

## **Why Does YPLL Matter?**

- YPLL is a measure of pre-mature mortality in all communities of all states.
- There is a multifaceted approach to staying healthy and they are all interconnected in many ways such as lack of education leads to economic disparity which in turn leads to not being able to access quality health care, quality living conditions which eventually leads to chronic health conditions and pre-mature mortality.
- YPLL is widely used metric to measure the health of the population by government in state and federal levels to monitor and hence take appropriate public health policy measures to improve the health of their population.

# Categories and Measures Closely Related to Health Outcomes



# Dataset Variables

## Health/ Behaviors/ Clinical Care

**Diabetes Prevalence** It is the percentage of adults aged 20 and above diagnosed with diabetes

**HIV rate** Rate of diagnosed cases of HIV for people aged 13 years and older in a county per 100,000 population

**Mental Health**  
**Provider Rate** - Ratio of the population to mental health providers

**Physical Inactivity**  
Percentage of adults ages 20 and reporting no leisure-time physical activity in the past month

## Physical Environments/ Social Factors

**Rural** Percentage of people living in rural areas of the county

**African American**  
Percentage of people that are African American

**65 and over** Percentage of people who are 65 years of age and older

**Female**  
Percentage of population that is female

**Drive Alone** - Percentage of the workforce that usually drives alone to work

## Economic Factors

**Median Household Income** Measure of income where half of households in a county earn more and half of households earn less

**High Housing Costs**  
Percentage of households with housing costs 30% or more of household income

**Free lunch** Percentage of children enrolled in public schools that are eligible for free lunch

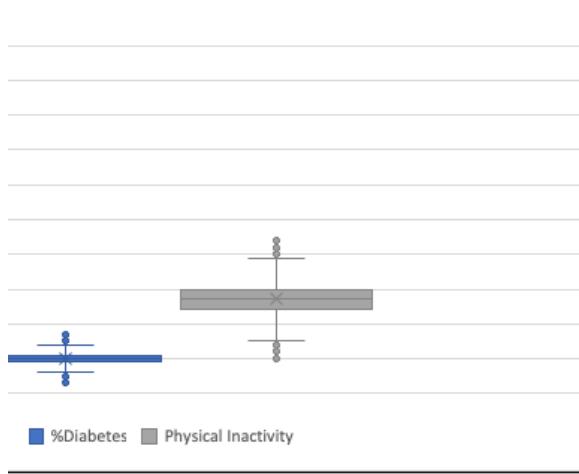
**% child Illiteracy**  
It is a measure of high school graduation rate

# Data Analysis

---

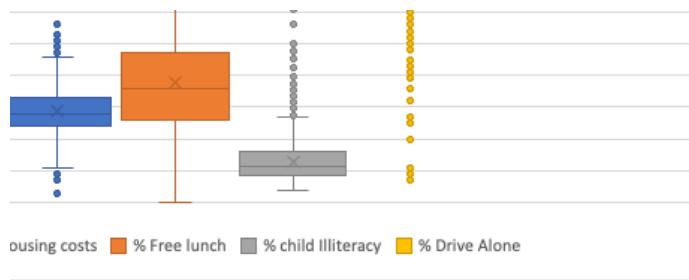
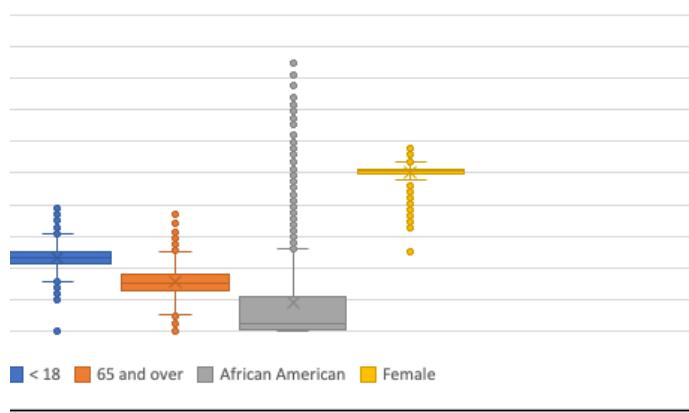
## Key Insights

- In 50% of the counties, less than 25% of the population reported physical inactivity
- Median percentage of population surveyed that have been diagnosed with diabetes health condition is 10%.
- Median household income data is high variability and several outliers that is positively skewed.
- The median income is about \$42,000. This also shows that there less people making more than \$70,000.

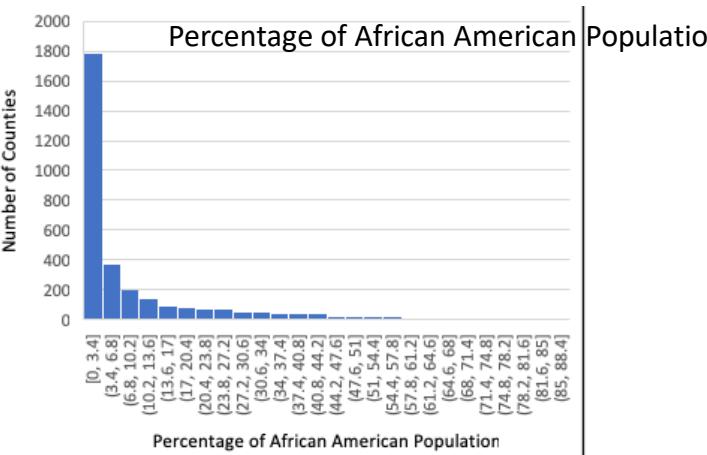


## Key Insights

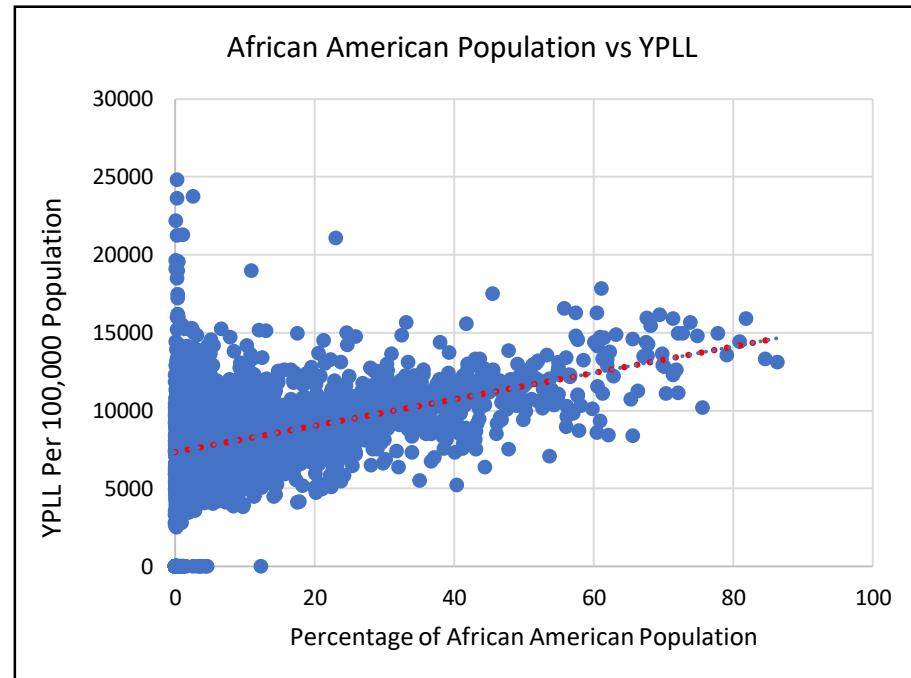
- Percentage of African American population in this study has several outliers and is positively skewed. The spread also shows high variability.
- More than 50% of African American population is concentrated in fewer counties while percentage of population that is females is more evenly distributed.
- Percentage of people driving alone is heavily negatively skewed.



**Scatter plot shows a positive correlation with YPLL for African American population in the counties in the US which points to increased mortality in African American population due to certain health conditions.**



African American	
Mean	9.193984962
Standard Error	0.253336877
Median	2.5
Mode	0.2
Standard Deviation	14.31297308
Sample Variance	204.8611985
Kurtosis	4.889730078
Skewness	2.223894974
Range	86.3
Minimum	0
Maximum	86.3
Sum	29347.2
Count	3192
Confidence Level(95.0%)	0.496719563

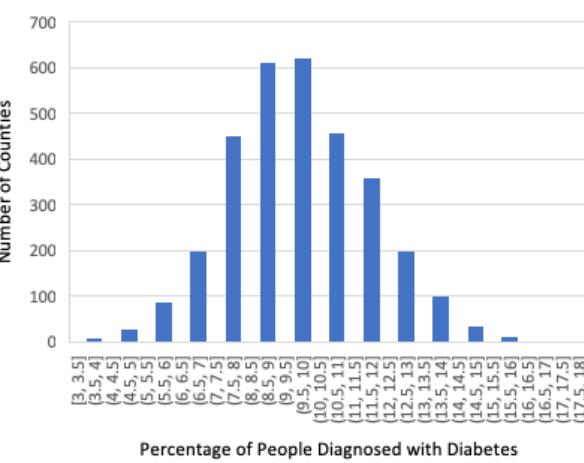


### Key Insights:

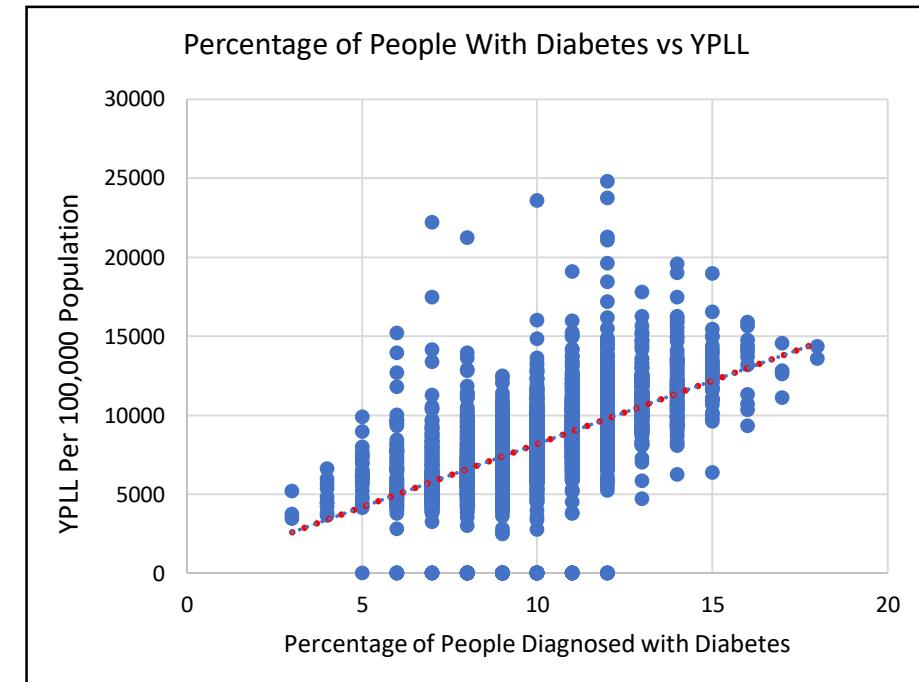
- % African American measures the percentage of population that is African American in all the counties in United States that were surveyed for this purpose. The numerator is the population of African Americans in all counties combined. The denominator is the total population in all counties combined.
- The histogram shows a long tail in the right indicating that the measure is highly positively skewed to the right and has several outliers. Also, about 50% of the African American population is in about 1800 counties surveyed.
- According to Health and Human Service (HHS) report on Minority group, death rates are higher in African American population due to cancer, 60% more likely to have diabetes and 20% more likely to die of heart disease. According to Center for Disease Control and Prevention (CDC), African Americans are more likely to die at younger ages due to these health reason. Most of this is attributed to social factors such as unemployment, living in poverty, obesity, lack of access to health care and inactivity.<sup>2</sup>

<sup>2</sup> <https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvid=16>

**Scatter plot shows a positive correlation between percentage of people diagnosed with diabetes and YPLL measure in the counties in US.**



%Diabetes	
Mean	9.935776942
Standard Error	0.036863959
Median	10
Mode	10
Standard Deviation	2.082732147
Sample Variance	4.337773197
Kurtosis	0.184837894
Skewness	0.187525023
Range	15
Minimum	3
Maximum	18
Sum	31715
Count	3192
Confidence Level(95.0%)	0.072279449

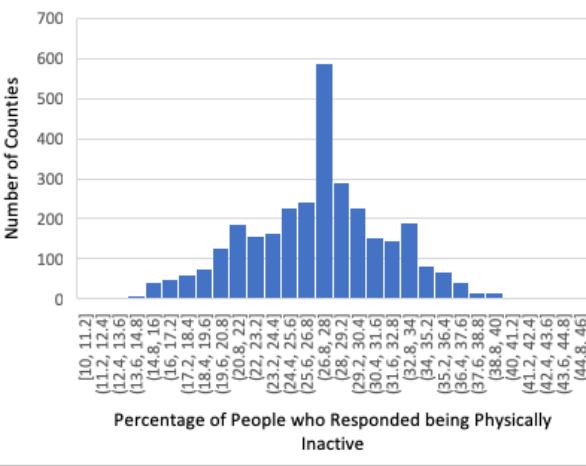


### Key Insights:

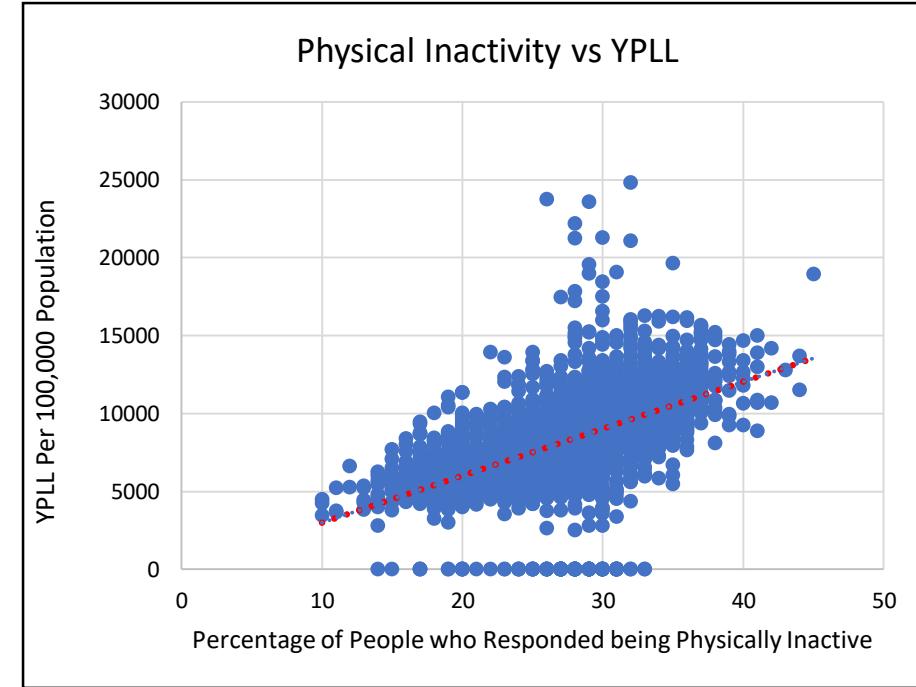
- Percentage of people with diabetes is the percentage of adults ages 20 and above with diagnosed diabetes (both Type 1 and Type 2) in all the counties in United States. The numerator is the number of adults 20 years and older who responded "yes" to the question, "Has a doctor ever told you that you have diabetes?" The denominator is the total number of respondents (age 20 and older) in a county.
- The histogram shows a normal distribution with a mean around 9.5%. The skewness represents approximately symmetric distribution.
- Approximately, 1200 counties have 8.5 to 10% of the population diagnosed with diabetes.
- According to CDC, diabetes ranks number 7 in mortality in the US which is in line with the data shown in the scatter plot.<sup>3</sup>

<sup>3</sup> <https://www.cdc.gov/nchs/fastats/diabetes.htm>

**Scatter plot shows positive correlation between percentage of adults of age 20 and over reporting no physical activity with YPLL in the counties in the US.**



Physical Inactivity	
Mean	27.05827068
Standard Error	0.089658579
Median	27
Mode	27
Standard Deviation	5.065511336
Sample Variance	25.65940509
Kurtosis	0.102365087
Skewness	-0.12862268
Range	35
Minimum	10
Maximum	45
Sum	86370
Count	3192
Confidence Level(95.0%)	0.175794265

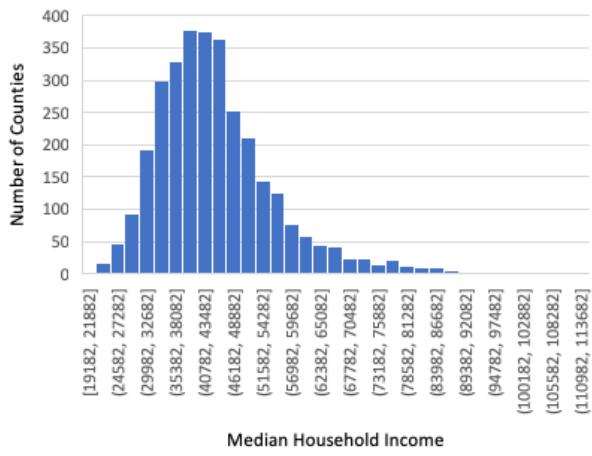


### Key Insights:

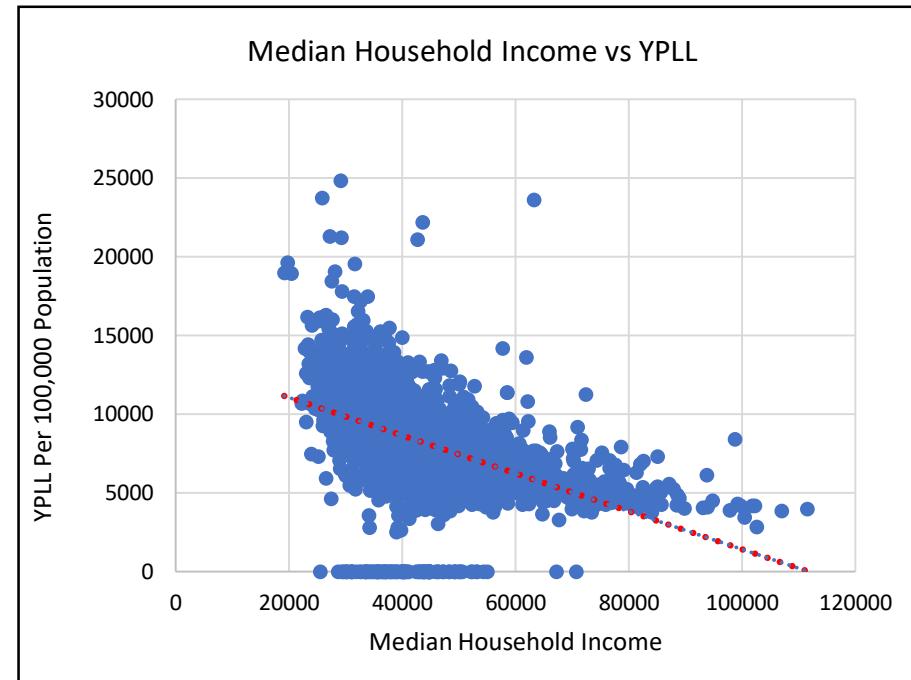
- Physical Inactivity is based on responses to the Behavioral Risk Factor Surveillance Survey (BRFSS) and is the percentage of adults ages 20 and over reporting no leisure-time physical activity in the past month.
- The numerator is the number of BRFSS Respondents who answered "no" to the question, "During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?" The denominator is the number of BRFSS respondents age 20 and older.
- The histogram shows almost normal distribution with slightly negative skewness. About 26 – 28% of the people who reported that they don't engage in physical activities are concentrated in about 600 counties.
- According to CDC and HHS, engaging in physical activity can help you age healthier and reduce risk of many chronic health conditions such as heart disease, diabetes and certain cancers.<sup>4</sup>

<sup>4</sup> <https://www.cdc.gov/physicalactivity/about-physical-activity/why-it-matters.html>

**Scatter plot shows negative correlation between median household income and YPLL in the counties in US indicating that low income has higher mortality rates.**



median household income	
Mean	44300.17612
Standard Error	202.8438651
Median	42554
Mode	39997
Standard Deviation	11458.43431
Sample Variance	131295716.8
Kurtosis	3.293576515
Skewness	1.386539962
Range	92400
Minimum	19182
Maximum	111582
Sum	141361862
Count	3191
Confidence Level(95.0%)	397.7175729

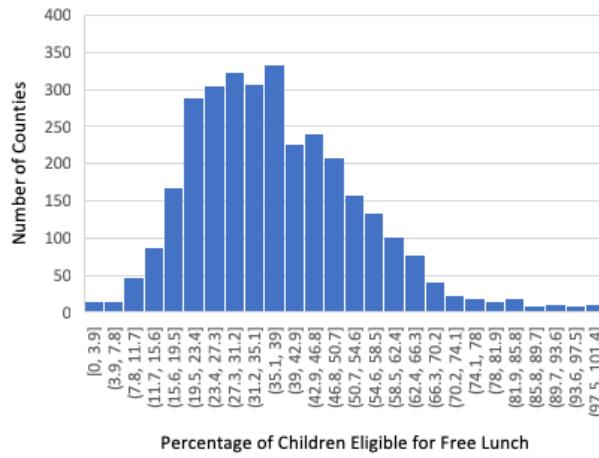


#### Key Insights:

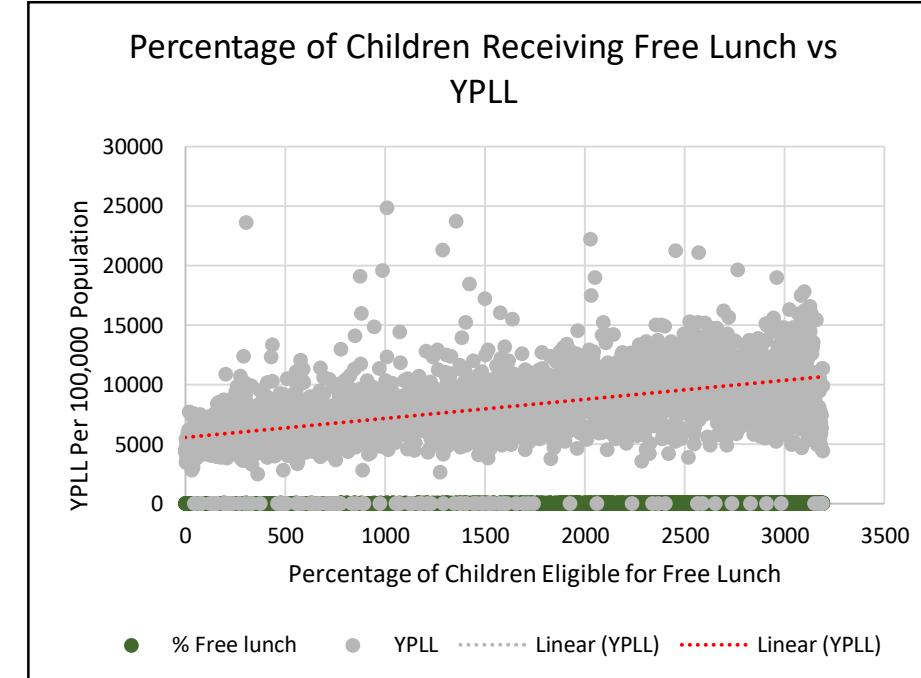
- Median Household Income is the income where half of households in a county earn more and half of households earn less.
- Median Household Income measure has a very low negative correlation with YPLL higher levels of mortality in house holds with lesser median household income.
- The histogram shows several outliers causing it to be heavily skewed to the right. There are fewer counties that have high median household income above \$62,000. Most common median income is between \$40,000 and 48,000 and that is spread in about 1110 counties.
- According to CDC, income is critical to get access to quality health care, quality education, nutrition good living conditions and more. These factors can have adverse effect on health and wellbeing of people. It can lead to serious and chronic health conditions, depression and suicide which results in high mortality rate thus increasing YPLL measure.<sup>5</sup>

<sup>5</sup> <https://www.cdc.gov/socialdeterminants/index.htm>

**Scatter plot shows positive correlation between percentage of children receiving free lunch in public schools and YPLL measure in the counties in the US.**



% Free lunch	
Mean	37.68799244
Standard Error	0.288856022
Median	36
Mode	38
Standard Deviation	16.27108306
Sample Variance	264.7481439
Kurtosis	0.83631075
Skewness	0.747620815
Range	100
Minimum	0
Maximum	100
Sum	119584
Count	3173
Confidence Level(95.0%)	0.56636351

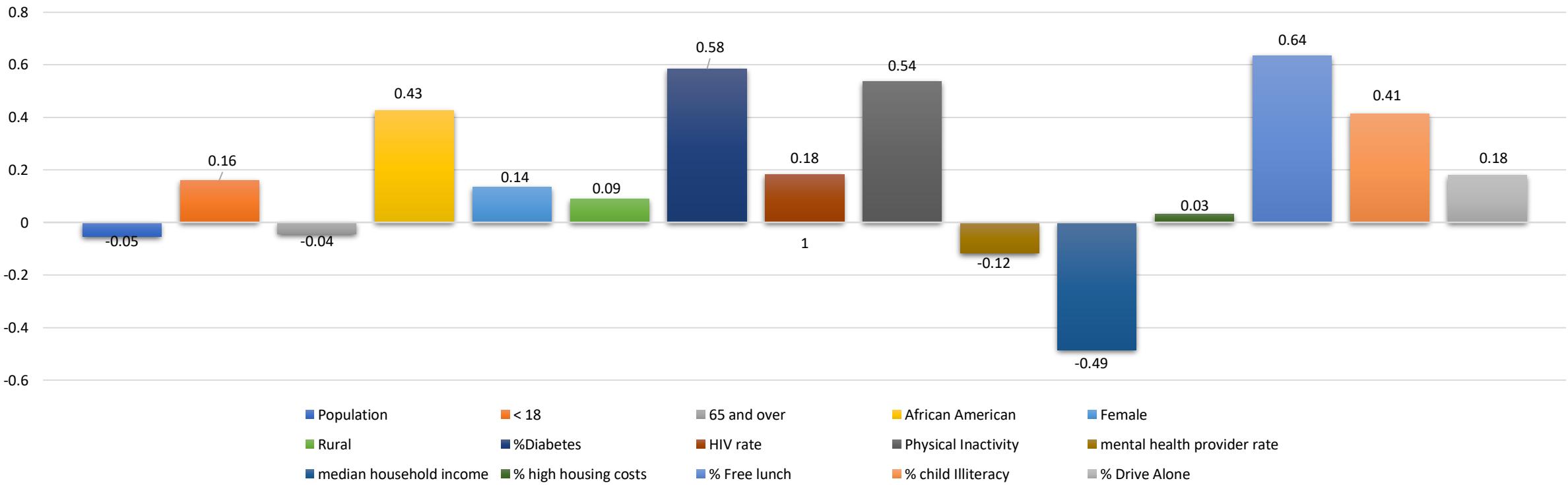


### Key Insights:

- Children Eligible for Free Lunch is the percentage of children enrolled in public schools that are eligible for free lunch.
- The numerator is the number of public school students, grades PK-12, eligible for free lunch. Children eligible for free lunch live in a family with income less than 130% of the federal poverty level.
- The denominator is the total number of students enrolled in public schools, grades PK-12.
- The histogram shows a almost normal distribution and moderately skewed to the right indicating several outliers. There are about 19% to 39% of children receiving free lunch are in 1500 counties of the country.
- According to <http://www.healthyfranklincounty.org/children-eligible-freereduced-price-lunch> "this indicator is relevant because it assesses vulnerable populations which are more likely to have multiple health access, health status, and social support needs." These families typically are poor and are vulnerable to chronic health conditions such as diabetes and heart disease. This could be a result of lack of access to quality healthcare, education and good living conditions. These issues can be addressed through good public health policies.

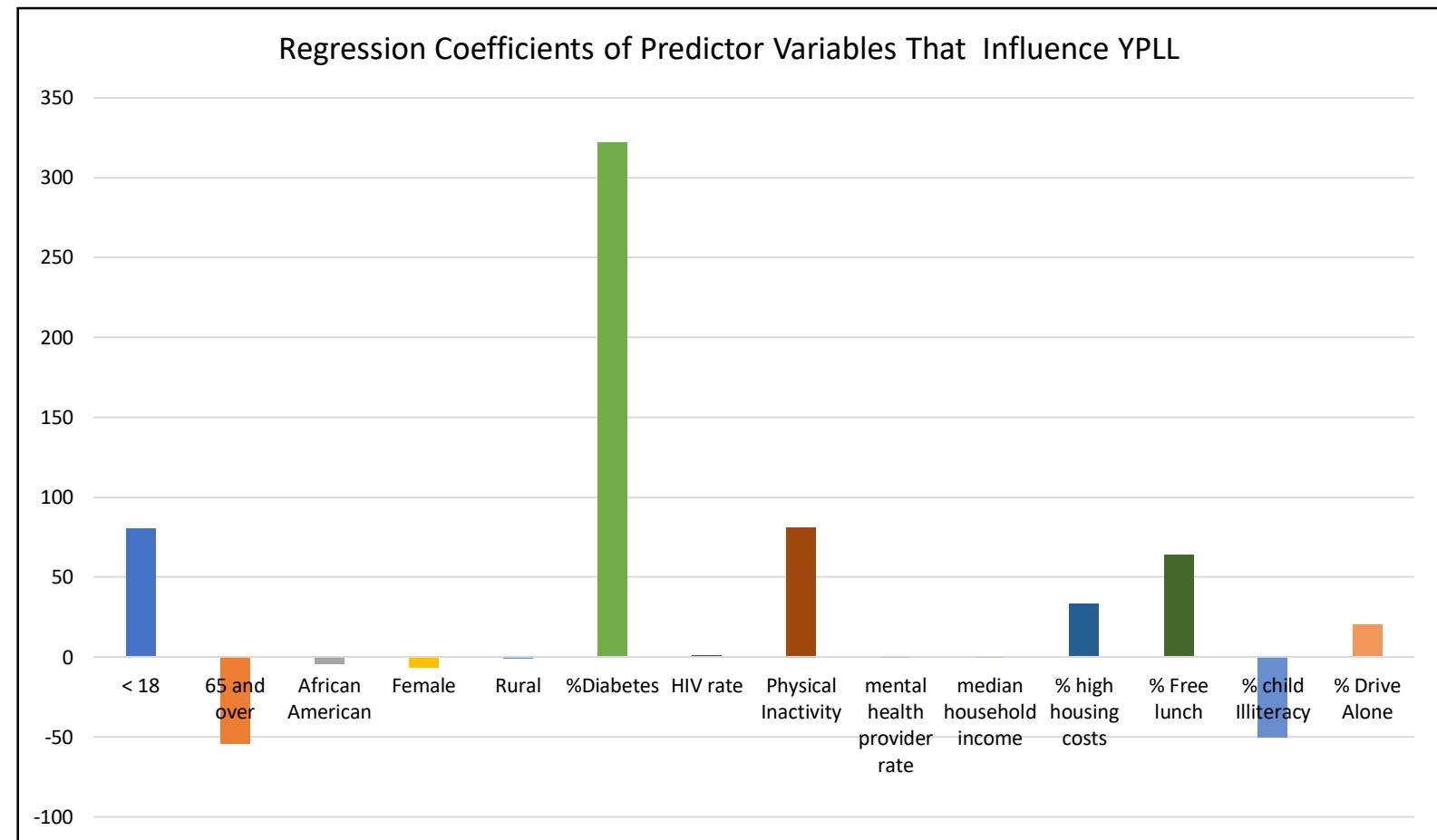
Three predictor variables that have high positive correlation with YPLL are percentage of children receiving free lunch, percentage of people diagnosed with diabetes and percentage of people who reported lack of physical activity. Also, important to note that median household income has a negative correlation with YPLL, indicating that poorer Americans have higher rates of mortality. This could be due to lack of access to quality healthcare, education and nutrition among other factors.

**Correlation Values of Variables Compared with YPLL**



**Percentage of diabetes diagnosed has the highest coefficient indicating that it has the highest impact on the YPLL while median household income of people in the counties has minimal impact on YPLL.**

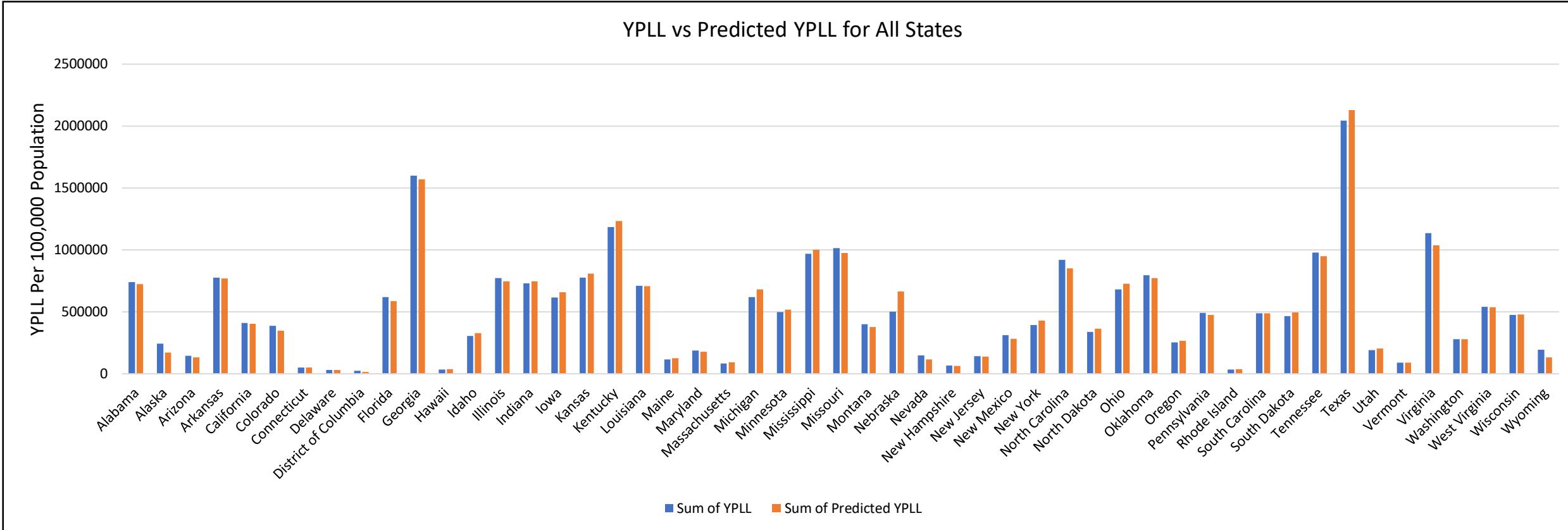
	Coefficients	P-value
Intercept	-539.23871	0.61337
Population	-5.612E-05	0.07358
< 18	79.9662096	0.00000
65 and over	-54.669563	0.00002
African American	-4.2836465	0.29391
Female	-6.5014913	0.74493
Rural	-1.0575932	0.49078
%Diabetes	321.764147	0.00000
HIV rate	0.9248834	0.00005
Physical Inactivity	81.1362121	0.00000
mental health provider rate	-0.3845991	0.53477
median household income	-0.0385063	0.00000
% high housing costs	33.1677793	0.00000
% Free lunch	63.7284058	0.00000
% child Illiteracy	-50.636425	0.00000
% Drive Alone	20.2742727	0.00012



#### Key Insights:

- Since we have several predictor variables, we got an adjusted R Square value of .507. That means 50% of the variance in YPLL can be explained by all the predictor variables.
- The P-Value indicates that percentage of African American population, percentage of female population and percentage of mental health providers are statistically insignificant ( $p > .05$ ) while all other predictor variables are statistically significant with a p-value  $< .05$ .

Regression model shows that YPLL Texas is the least healthiest state and District of Columbia is the most healthiest state.



#### Key Insights:

- Regression model shows that the measured YPLL vs Predicted YPLL is in line for most of the states.
- There is 4% increase in predicted YPLL for Texas and 32% increase in predicted YPLL in counties in Nebraska.
- There is 7% decrease in predicted YPLL for counties in North Carolina and 8% decrease in counties in Virginia.
- Indiana, Iowa, Kentucky, Michigan, Nebraska, Ohio and Texas seem to have predicted YPLL higher than the measured YPLL.
- Colorado, Florida, Georgia, Missouri, New Mexico, Oklahoma, Pennsylvania, Tennessee, Virginia and West Virginia have predicted YPLL lower than the measured YPLL.

# Findings & Implications

## Who is at risk?

---

Everyone is at risk but most at risk are Black/African American, Hispanic/Latino, people. They are disproportionately poorer and systemically discriminated and consequently at higher risk of pre-mature mortality.

Women due to educational and economic disadvantages

Elderly population often suffering from food insecurity and chronic health problems due to disability, mobility issues and physical inactivity among others

Children and Young adults due to illiteracy, abuse, depression, drug-overdose, and motor vehicle accidents and more

- There are major implications to communities that are plagued with health, social and economic issues that lead to pre-mature mortality. First of all, it is tragedy that people lose their lives who otherwise could have been happy and successful members of the community. Each and everyone matter and their contribution to the society matters.
- If people in the community are not healthy, it costs the government money to provide health care for the uninsured, increase in Medicare and Medicaid costs.
- If people in the communities do not have access to education, it has adverse effect on their economic status, leads to poverty, unable to find sustainable jobs, could lead to depression, violence and poor mental health conditions, all of which can lead to pre-mature mortality.
- According to Europe PMC1, “higher levels of incarceration are associated with higher levels of both morbidity (percentage reporting fair or poor health) and mortality (life expectancy)” in the communities. This is not only a security risk to the population in the communities but also costs the government huge amounts of money to keep criminals locked.

# Limitations

---

## Limitations

- It is important to note that regression models shows that only 50% of the variance in YPLL can be explained by all the predictor variables considered in this study combined.
- This study only includes limited number of variables that can influence morbidity and mortality of the population.
- It is also important to note that there is a high collinearity issue with some of the variables in this dataset. For instance, variables like African American population in communities are closely related to poverty, illiteracy and HIV rate which can increase the variance of the coefficient estimates.

# Who Can Help?

---

# Role of the Government

Policy makers passing laws and public health policies that provides healthcare for all and increase he minimum wages that can help families live a healthier life to name a few.

Government providing funding for building infrastructure like schools, colleges and hospitals

Government providing funding to make sure the safety and quality of food and water is regulated

Government providing funding and resources to organizations conducting biomedical research.

Local government and community leaders organizing community events to educate and bring awareness to people about health risks related to smoking, AIDS, drug overdose, motor vehicle accidents and more

Providing affordable public housing for people

# Role of Public and Private Sector Institutions

Educational institutions offer quality educational opportunities to people regardless their economic status.

Corporations providing job opportunities to people

Medical communities providing quality care for people in all communities

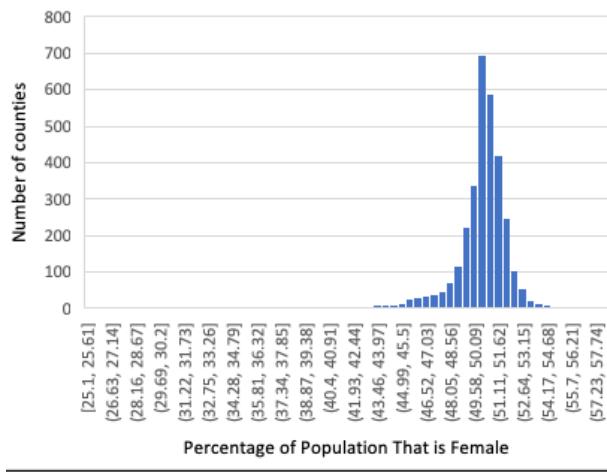
People having opportunities to pre-screen for certain chronic health conditions as preventive care access to necessary vaccinations.

People having access to mental care providers and community youth support programs more.

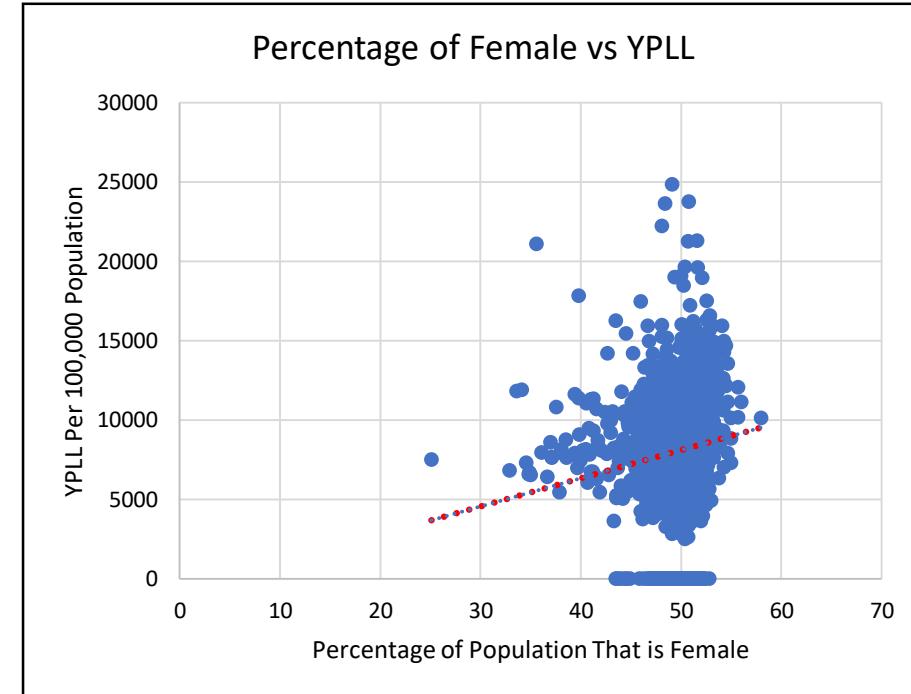
# Appendix

---

## Scatter plot shows positive correlation between percentage of female population mortality and YPLL measure in the counties in the US.



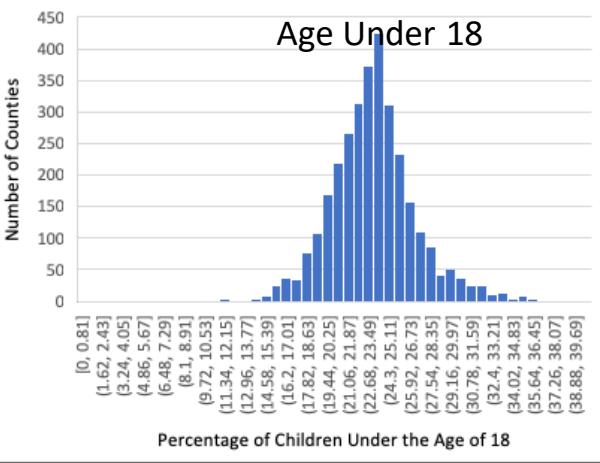
Female	
Mean	50.17807018
Standard Error	0.038333853
Median	50.5
Mode	50.5
Standard Deviation	2.165777881
Sample Variance	4.69059383
Kurtosis	19.03055589
Skewness	-3.263633215
Range	32.9
Minimum	25.1
Maximum	58
Sum	160168.4
Count	3192
Confidence Level(95.0%)	0.07516148



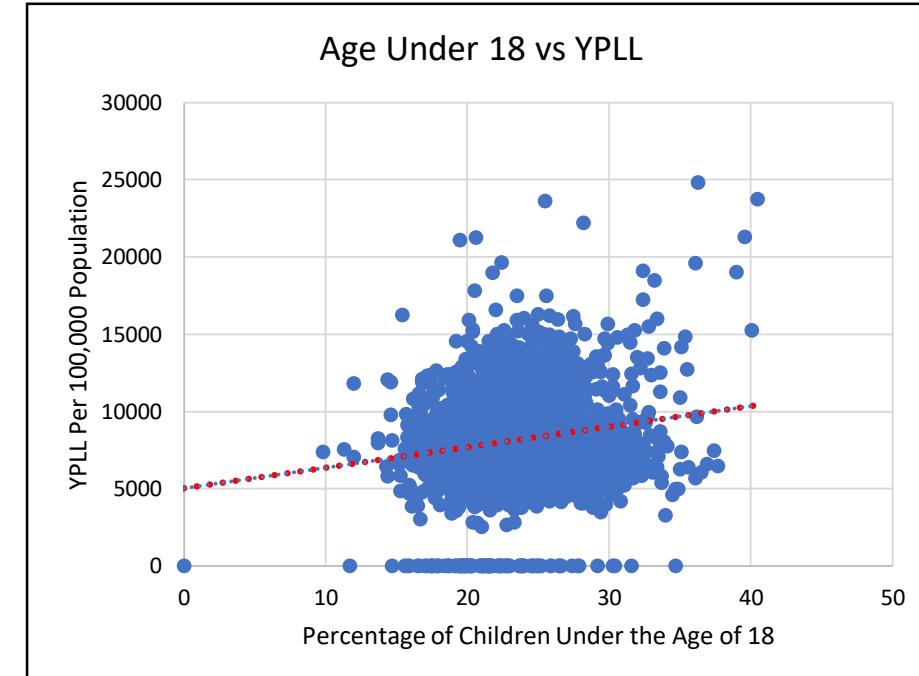
### Key Insights:

- Percentage of females measures the percentage of the population that is female.
- The numerator is the female population in the counties combined.
- The denominator is the total population in the counties combined.
- The histogram shows a almost normal distribution and is highly skewed to the left indicating several outliers.
- Even though mortality rates in males is higher than female, several chronic diseases that are prevalent in me are also affection women. Some of them can be addressed by providing quality heath care access to women so that they can pre-screen for certain health conditions as preventive care measures.

## Scatter plot shows positive correlation between number of children under 18 mortality and YPLL measure in the counties in the US.



Population Under the Age of 18	
Mean	23.47703634
Standard Error	0.060356313
Median	23.4
Mode	23.5
Standard Deviation	3.409998147
Sample Variance	11.62808736
Kurtosis	2.330701259
Skewness	0.447921058
Range	40.5
Minimum	0
Maximum	40.5
Sum	74938.7
Count	3192
Confidence Level(95.0%)	0.118341087

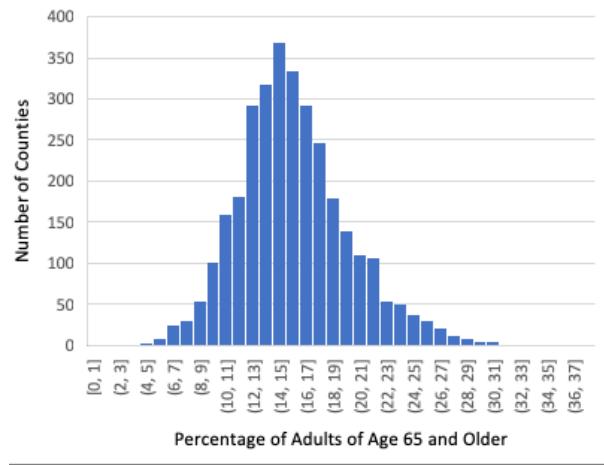


### Key Insights:

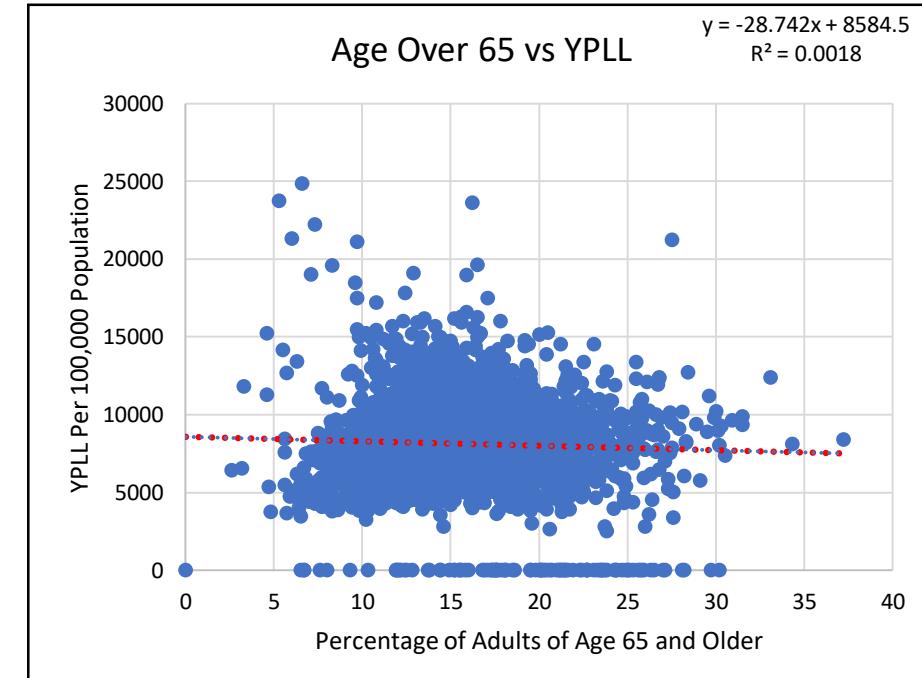
- This is a measure of mortality rate in children under the age of 18 years.
- The numerator is the number of children who dies under the age of 18.
- The denominator is the total population under the age of 18.
- The histogram shows a almost normal distribution and is approximately skewed.
- According to CDC, even though deaths among teenagers make up a small fraction of all deaths, some of them are preventable deaths because the most common are death due to motor vehicle accidents, homicide, suicide, drug-overdose and other chronic health conditions. Some of these stem from poverty and lack of opportunities.<sup>6</sup>

<sup>6</sup><https://www.cdc.gov/nchs/products/databriefs/db37.htm>

**Scatter plot shows slightly negative correlation between percentage of population 65 years of age and older and YPLL measure in the counties in the US.**



65 and over	
Mean	15.59263784
Standard Error	0.074616937
Median	15.2
Mode	14.3
Standard Deviation	4.215691873
Sample Variance	17.77205797
Kurtosis	0.947365401
Skewness	0.574041472
Range	37.2
Minimum	0
Maximum	37.2
Sum	49771.7
Count	3192
Confidence Level(95.0%)	0.146302003

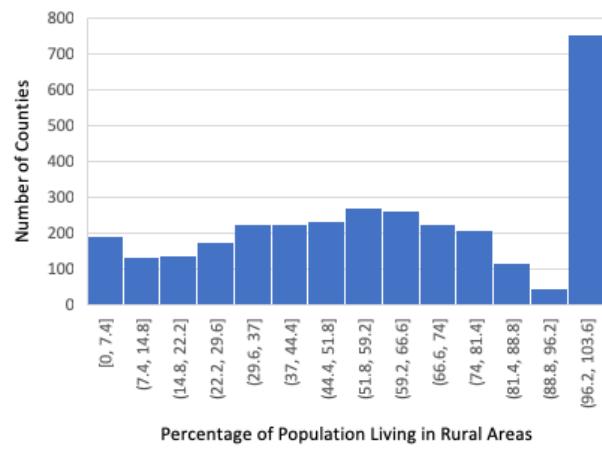


### Key Insights:

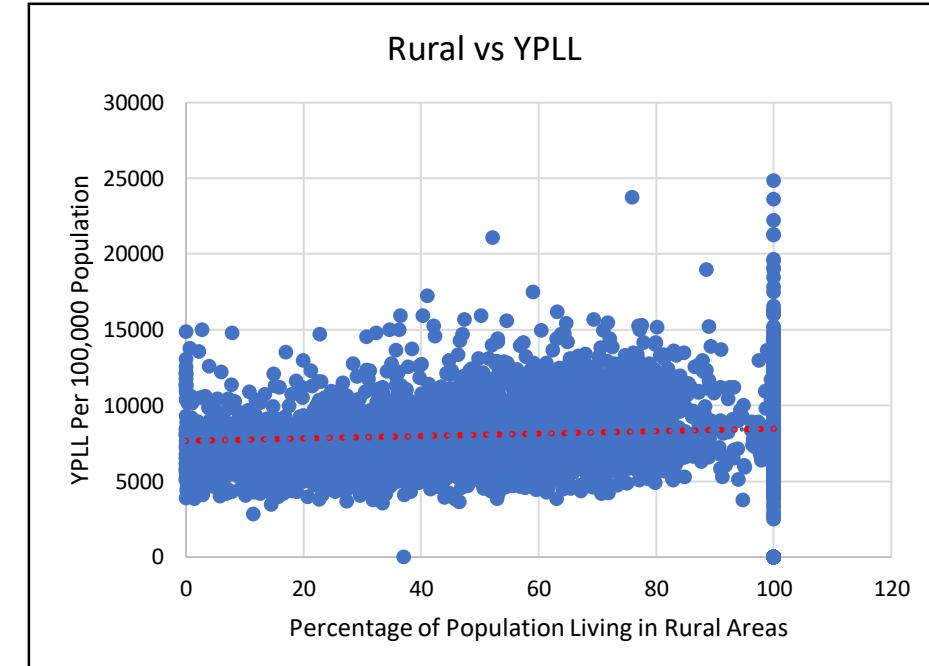
- This is a measure of mortality rates among population in US counties that are of age 65 years and older.
- The numerator is the number of people aged 65 years and older in all counties combined in the US.
- The denominator is the total population in all counties in the US combined.
- The histogram shows a almost normal distribution and moderately skewed to the right indicating several outliers.
- The negative correlation between percentage of population that are 65 years of age and older and YPLL.
- According to americashealthrankings.org <sup>7</sup>, in spite of improved health care practices geared toward seniors like preventive health care screening, pre-mature death is still a cause of concern among adults between ages 65 to 74. This is mainly contributed to food insecurity and obesity facing the elderly population because of their in-ability to go get food. Again, this can be preventable through public health policy changes and community activities.

<sup>7</sup> <https://www.americashealthrankings.org/learn/reports/2017-senior-report/findings-national-findings>

**Scatter plot shows slight positive correlation between mortality rate in population living in rural areas and YPLL measure in the counties in the US.**



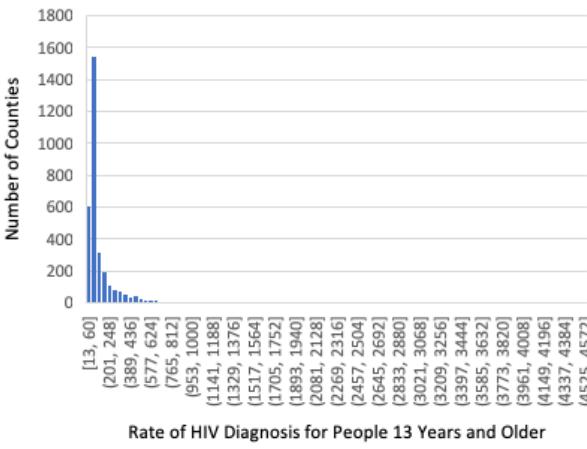
Rural	
Mean	59.38451896
Standard Error	0.549660868
Median	59.6
Mode	100
Standard Deviation	31.0497581
Sample Variance	964.0874781
Kurtosis	-1.108387889
Skewness	-0.166709608
Range	100
Minimum	0
Maximum	100
Sum	189496
Count	3191
Confidence Level(95.0%)	1.077724417



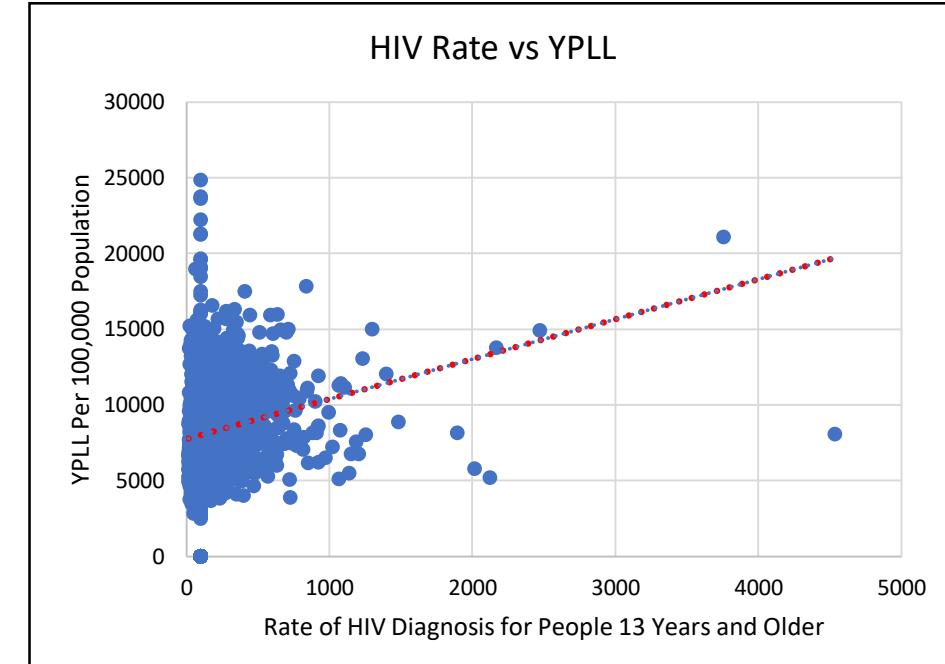
### Key Insights:

- This is a measure of percentage of population living in Rural counties in the US.
- The numerator is the population living in areas in counties classified as rural.
- The denominator is the total population in all counties combined in the US.
- The histogram shows a almost normal distribution and moderately skewed to the left indicating several outliers.
- Mortality rate of population in rural areas of US is not very significant.
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4590732/>

## Scatter plot shows positive correlation between HIV rates and YPLL measure in the counties in the US.



HIV rate	
Mean	170.3372197
Standard Error	4.89568245
Median	99
Mode	43
Standard Deviation	231.1882043
Sample Variance	53447.98583
Kurtosis	96.79099778
Skewness	7.248225193
Range	4521
Minimum	13
Maximum	4534
Sum	379852
Count	2230
Confidence Level(95.0%)	9.600574421

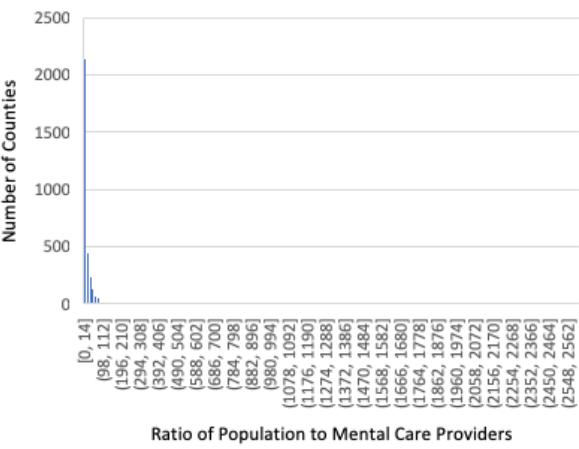


### Key Insights:

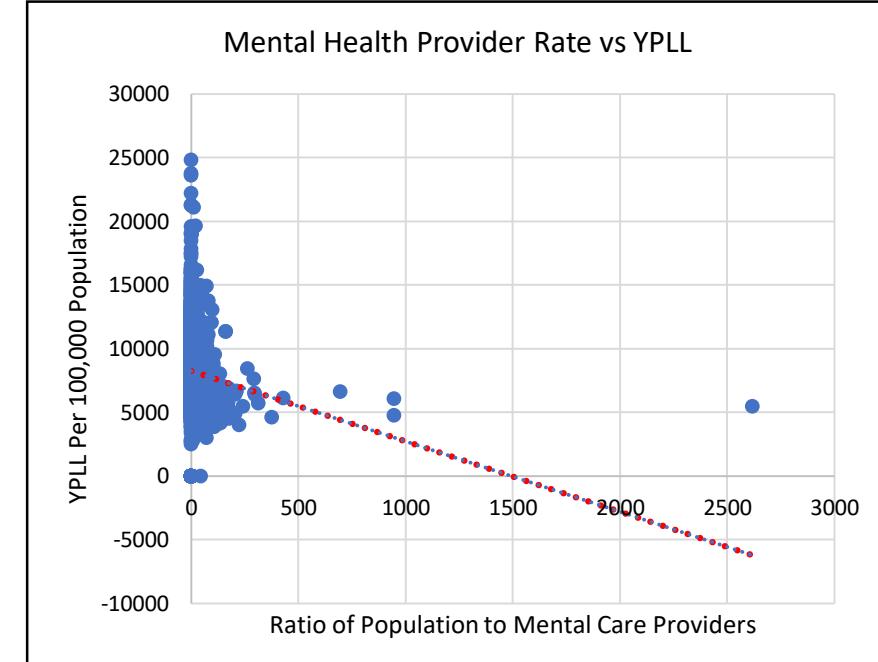
- HIV Prevalence is the rate of diagnosed cases of HIV for people aged 13 years and older in a county per 100,000 population.
- The numerator is the number of HIV cases in all the counties combined.
- The denominator is the total population in all counties combined.
- The histogram shows a almost normal distribution and highly positively skewed to the right indicating several outliers.
- According to CDC <sup>9</sup>, HIV infections have been reducing in all the states in US. However, it is still a major cause of concern because according to CDC statistics, the diagnosis is highest among younger Americans between the age of 23-34. Also, it disproportionately affects black/African Americans and Hispanic/Latino population who are already vulnerable demographic population.

<sup>9</sup> <https://www.cdc.gov/hiv/statistics/overview/ataglance.html>

**Scatter plot shows negative correlation between mental health provider ratio and YPLL measure in the counties in the US indicating higher pre-mature mortality in counties that have lesser access to mental health providers.**



mental health provider rate	
Mean	18.10588972
Standard Error	1.072265051
Median	6
Mode	0
Standard Deviation	60.58060306
Sample Variance	3670.009467
Kurtosis	1103.997885
Skewness	28.01569238
Range	2618
Minimum	0
Maximum	2618
Sum	57794
Count	3192
Confidence Level(95.0%)	2.102398328



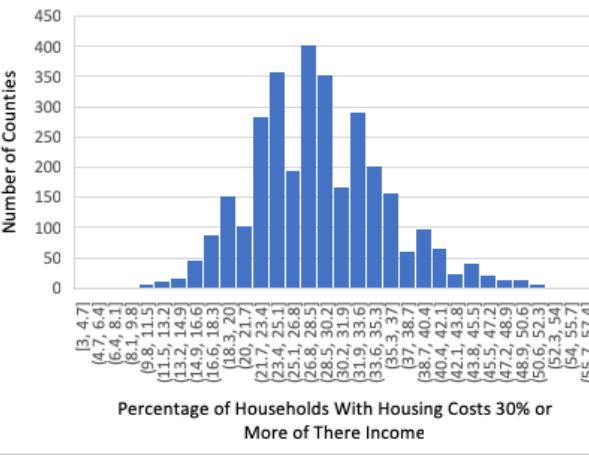
#### Key Insights:

- Mental health providers is a ratio of population to mental care providers in counties, if the population were equally distributed across providers.
- The left side of the ratio is the population in all counties combined.
- The right side of the ratio is the number of mental health providers in all counties combined.
- The histogram shows that the data is highly positively skewed to the right indicating several outliers. It also shows that there are more counties with less mental health provider to population ratio.
- Even though this measure is less significant compared to other measures that contribute to YPLL, according CDC <sup>10, 11</sup>, there is gap between people suffering with some type of mental disorder and the mental health provider accentuating the need for more mental health providers. Once again, like many other measures, access to mental health providers is rural area is difficult.

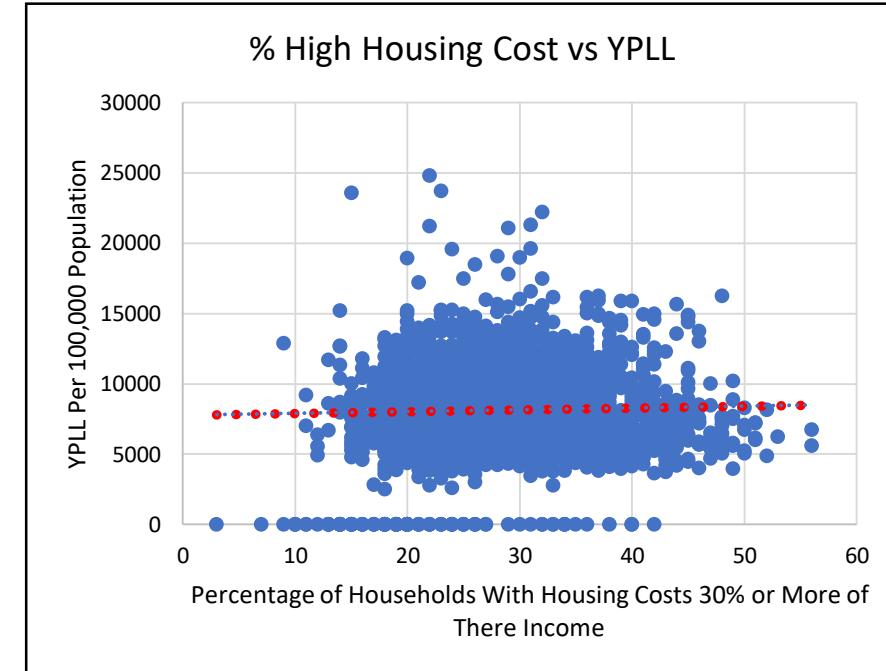
<sup>10</sup> <https://www.cdc.gov/ruralhealth/child-health/policybrief.html>

<sup>11</sup> <https://www.americashealthrankings.org/explore/annual/measure/MHP>

**Scatter plot shows almost negligible correlation between percentage of housing costs and YPLL measure in the counties in the US indicating positive correlation between housing insecurity and pre-mature mortality.**



% high housing costs	
Mean	28.72775689
Standard Error	0.123223047
Median	28
Mode	28
Standard Deviation	6.961829509
Sample Variance	48.46707012
Kurtosis	0.376046602
Skewness	0.413560114
Range	53
Minimum	3
Maximum	56
Sum	91699
Count	3192
Confidence Level(95.0%)	0.241604375

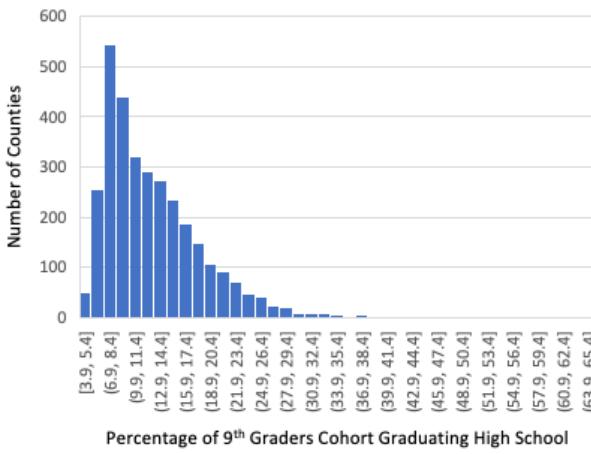


### Key Insights:

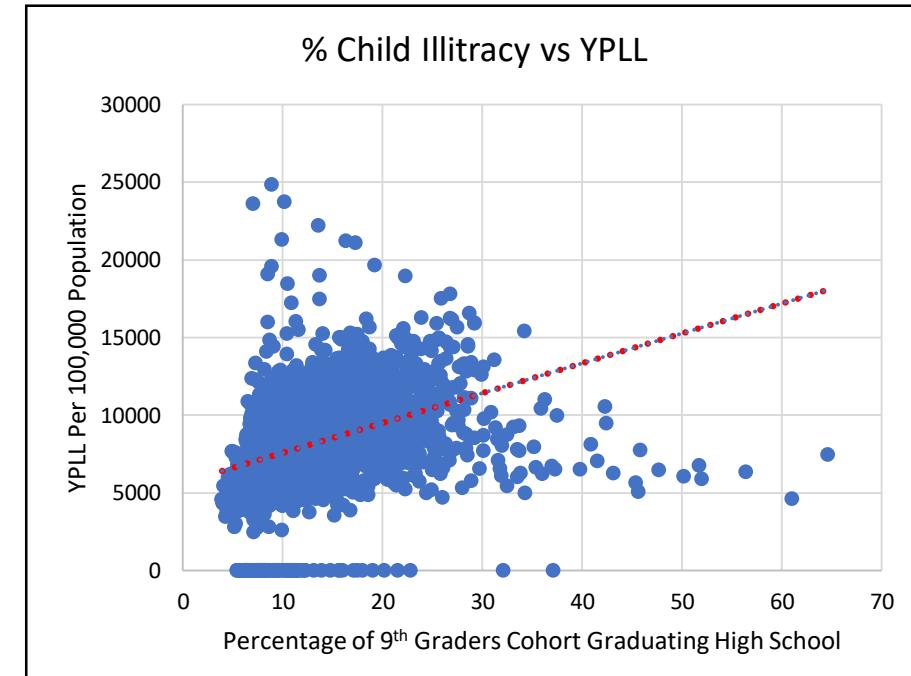
- High Housing Costs represents the percentage of households with housing costs 30% or more of household income.
- The numerator is the number of renter-occupied housing units or owner-occupied housing units with a mortgage that pay 30 percent or more of their household income on housing costs in the county.
- The denominator is the total number of housing units in the county.
- The histogram shows a almost normal distribution and the skewness shows that the distribution is approximately symmetric.
- According to National Institute of Health (NIH)<sup>12</sup>, "Poor housing conditions are associated with a wide range of health conditions, including respiratory infections, asthma, lead poisoning, injuries, and mental health.". According to the same article, there is a positive correlation between quality of housing and chronic illnesses that can eventually lead to pre mature mortality.

<sup>12</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447157/>

**Scatter plot shows positive correlation between child illiteracy and YPLL measure in the counties in the US indicating how lack of education can have serious health consequences leading to pre-mature mortality.**



% child Illiteracy	
Mean	12.86610169
Standard Error	0.108802945
Median	11.4
Mode	7.6
Standard Deviation	6.141345559
Sample Variance	37.71612527
Kurtosis	6.903777591
Skewness	1.871433322
Range	60.7
Minimum	3.9
Maximum	64.6
Sum	40991.4
Count	3186
Confidence Level(95.0%)	0.213330922



### Key Insights:

- High School Graduation is the percentage of the ninth-grade cohort that graduates from high school in four years.
- The numerator is the number of cohort members who earned a regular high school diploma by the end of the school year.
- The denominator is the number of first-time 9th graders four years prior to graduation, plus students who transferred in, minus students who transferred out, emigrated, or died during school years.
- The histogram shows a almost normal distribution and highly skewed to the right indicating several outliers.
- According to [worldliteracyfoundation.org](https://worldliteracyfoundation.org)<sup>13</sup>, lack of education has multifaceted consequences in the communities and it's population. It causes social problems which mostly affects poorer people and girls, economic impact because, acquiring jobs is directly linked to education, Health consequences which is severely impacted if people cannot access or afford quality health care and make poor behavioral and health choices, to name a few.

<sup>13</sup> <https://worldliteracyfoundation.org/>



# What is Age-Adjusted Rate?

- Age adjusted rates are used to address any bias that can be introduced when measuring death rates in population of different age groups. This is because, some counties can have more elderly people compared to younger people. This means, the death rates will be higher in that county and becomes difficult to compare death rates among such populations.
- In order to make it fair, the age distribution of the population is adjusted to a “standard” population distribution.

# How is “Age-Adjusted Rate” calculated?

According to Missouri Department of Health & Senior Services <sup>14</sup>, it is computed using a formula called “direct method of standardization” that is described below:

- The rate in the area of study (e.g., county, state) is computed for each age group noted in “U.S. Standard Population Distribution” table <sup>14</sup> by dividing the number of events (deaths) in that age group by the estimated population of the same age group in that area and then multiplying by a constant of 100,000.
- This results in an age-specific death rate (ASDR) per 100,000 population for each age group. That is, for each age group,  $\text{ASDR} = \text{deaths in age group} \div \text{estimated population of that age group} \times 100,000$ .
- Each ASDR is then multiplied by the proportion of the standard population table <sup>14</sup>, that same age group to get the age-adjusted death rate (AADR). The age-specific results are summed to get the age-adjusted death rate for the area of study. The formula is:
  - $\text{AADR} = \text{Summation of (ASDR} \times \text{standard proportion)}$

<sup>14</sup> [https://health.mo.gov/data/mica/CDP\\_MICA/AARate.html](https://health.mo.gov/data/mica/CDP_MICA/AARate.html)