Hayden Schindler  
5014 Westway Dr, Apt. 714

San Angelo, TX 76904  
+1 3252458757

hschindler1@angelo.edu

**ABSTRACT:**

This data science project utilized the UHG 2022 census data for the United States to investigate several potential correlations between various socioeconomic factors and important health and social outcomes. Specifically, the project aimed to answer the research questions below:

# INTRODUCTION:

The data for the project was obtained from the county health rankings dataset for 2022, which can be found at: (<https://www.countyhealthrankings.org/sites/default/files/media/document/2022%20county%20health%20rankings%20data%20-%20v1.xlsx>.)

The project used the county health rankings dataset for 2022 as the primary data source, and the research method involved isolating the relevant information and drawing conclusions based on the findings. Overall, this project highlights the complex relationships between socioeconomic factors and health outcomes and underscores the importance of continued research in this field. The project employs data analysis techniques such as correlation analysis and visualization to draw insights from the dataset.

**RESEARCH QUESTIONS:**

* Is there a relationship between areas with higher broadband access and higher suicide rates or more poor mental health days?
* Does a higher number of single-parent households in communities correspond to higher rates of teen pregnancy?
* Are there connections between areas with higher rates of insufficient sleep and lower graduation rates or higher rates of disconnected youth?
* Can a connection be established between areas with a higher number of physically inactive days and limited access to healthy foods or frequent physical distress?
* Is there any correlation between higher suicide rates and areas that score lower for home ownership?

# RELATED WORK & RESULTS:

**Q1:**

There appears to be no distinct correlation between broadband access and suicide rate. The states with the four highest suicide rates (Wyoming, New Mexico, New York, and California), all have different rates of broadband access, and on top of that, the broadband access in every state is high to the point where it doesn't seem like any small dip of a percent has a major effect.

**Heat map for Q1:**

Graphical user interface, website

Description automatically generated

**Correlation chart for each category:**

**Diagram, engineering drawing

Description automatically generated**

**Q2:**

It appears that there is some correlation between single parent households and teen birth rate. Specifically, looking at states with higher teen birth rates (Mississippi, Oklahoma, etc.) They tend to have high/midrange single parent household rates. When it comes to single parent households and arrest rates, the results are a bit inconclusive due to a lack of stats provided by the dataset, but it appears there is some correlation there from the data that is there. States with lower juvenile arrest rates tend to be in the low/mid-range for single parent households, whereas states with high juvenile arrest rates tend to be in the mid to high range for single parent households.

**Heat map for Q2:**

Graphical user interface, text

Description automatically generated

**Correlation chart for each category:**

**Chart

Description automatically generated**

**Q3:**

There isn't any direct correlation between the percentage of insufficient sleep and the high school graduation rate because the areas that reported the highest level of sleep inefficiency, also have the highest graduation rate in the country, but the graduation rate is high for majority of the country. There does appear to be some correlation between sleep inefficiency and the percentage of disconnected youth. This can mainly be shown in counties in the states of New Mexico, Nevada, Oklahoma, Texas, Florida, and Michigan.

**Heat map for Q3:**

Graphical user interface, website

Description automatically generated

**Correlation chart for each category:**

**Chart, scatter chart

Description automatically generated**

**Q4:**

There didn't appear to be a correlation between access to healthy food and either of the two categories. But there did appear to be a direct correlation between the percentage of physically inactive people and the percentage of being frequently distressed to the point where the maps nearly look identical. This is most prominently shown in the states of Texas, Mississippi, Oklahoma, Georgia, and West Virginia.

**Heat map for Q4:**

Graphical user interface, application

Description automatically generated

**Correlation chart for each category:**

**Chart

Description automatically generated**

**Q5:**

Like question one, where it might be assumed that lower homeowner rate would lead to higher suicide rates, based on the bar graphs it appears that would be false. Looking at the lowest homeowner percent (DoC), they also have the lowest rate of suicide. And even places with much higher homeowner percentages like West Virginia or Louisiana still have a high suicide rate, almost twice so compared to the District of Columbia. The states with the highest suicide rates (Wyoming, New Mexico, New York, and California) all also have decent homeownership precents.

**Heat map for Q5:**

Graphical user interface

Description automatically generated with low confidence

**Correlation chart for each category:**

Chart, line chart, scatter chart

Description automatically generated

# Data sets:

Using the UHG census data in its entirety to form these connections but the data was consolidated down to the following categories to make more concise connections. The Social & Economic category accounted for 40% of the health factors that contribute to overall health outcomes. This is why the data set was condensed to look at the individual contributors to the Social & Economic factors. Below the individual factors are listed to provide context for some of the diagrams and relationships.

## QUALITY OF LIFE FACTORS

* % Fair or Poor Health
* Average Number of Physically Unhealthy Days
* Average Number of Mentally Unhealthy Days
* % Low birthweight
* % Frequent Physical Distress
* % Frequent Mental Distress
* % Adults with Diabetes
* HIV Prevalence Rate

## LENGTH OF LIFE FACTORS

* Years of Potential Life Lost Rate
* Life Expectancy
* Age-adjusted Death Rate
* Child Mortality Rate
* Infant Mortality Rate

## SOCIAL AND ECONOMIC FACTORS

* % Completed High School
* % Some College
* High School Graduation Rate
* % Disconnected Youth
* Average Grade Performance (Reading)
* Segregation index
* Spending per-pupil
* % Unemployed
* % Children in Poverty
* Gender Pay Gap
* Median Household Income
* % Enrolled in Free or Reduced Lunch
* % Children in Single-Parent Households
* Social Association Rate
* % household income required for childcare expenses
* Injury Death Rate
* Homicide Rate
* Suicide Rate (Age-Adjusted)
* Firearm Fatalities Rate
* Motor Vehicle Mortality Rate
* Juvenile Arrest Rate

## PREDICITVE MODELING INTRODUCTION FOR DATA SETS

Social & Economic Factors make up 40% of distribution relating to health outcomes so It was decided that would be a good place to start. The individual factors of each category were broken down and leveraged against each other to show how each category contributes to the whole, that is Social & Economic factors. The highest correlating factors of the S&EF were % Frequent physical distress and the average number of physically unhealthy days and the factors of Average number of mentally unhealthy days and the % Frequent mental distress. Upon consolidating, there appears to be a correlation between % children in poverty and the % Frequent physical distress, which contributes to negative health outcomes. Upon inspecting factors through correlation maps, the % Children in Poverty seemed like a place to explore as it could be the potential root cause of some lower ranging scores for the categories contained in the Social & Economic factors and it could subsequently be a key proponent of some of the negative health behaviors and health outcomes. The first categories subjected to predictions were the rate of children in poverty and the rate of frequent physical distress.

## PREDICITVE MODELING ANALYSIS

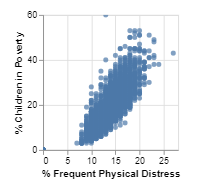
The first model created aimed to explore the association between the rate of children living in poverty and their frequency of physical distress. Through the use of a heat map, the initial investigation revealed that the two variables were highly correlated. In other words, areas with a higher percentage of children living in poverty tended to have higher rates of physical distress among children as well. Upon further analysis, the predictive model confirmed these results, indicating that a higher prevalence of child poverty was a strong predictor of increased physical distress. These findings are significant for several reasons. Firstly, higher rates of physical distress are associated with a range of negative health outcomes, including a greater likelihood of chronic conditions, reduced quality of life, and decreased life expectancy. Additionally, physical distress may also lead to the development of negative health behaviors, such as smoking, drinking, or drug abuse, which further exacerbate negative health outcomes. Furthermore, children living in poverty may experience a range of adverse health outcomes beyond physical distress. For example, they may be at an increased risk of poor mental health, malnutrition, or developmental delays, all of which can have long-lasting effects on their overall health and wellbeing. Given these consequences, it is imperative that we prioritize efforts to reduce child poverty and alleviate its associated health burdens. Effective strategies may include increasing access to affordable housing, improving education and job training opportunities, and implementing policies that ensure all families have access to quality healthcare. By taking these steps, we can promote better health outcomes for children, reduce the incidence of physical distress and its negative consequences, and create a healthier and more equitable society.

# PREDICTIVE MODELS

## children in poverty rate & frequent physical distress rate

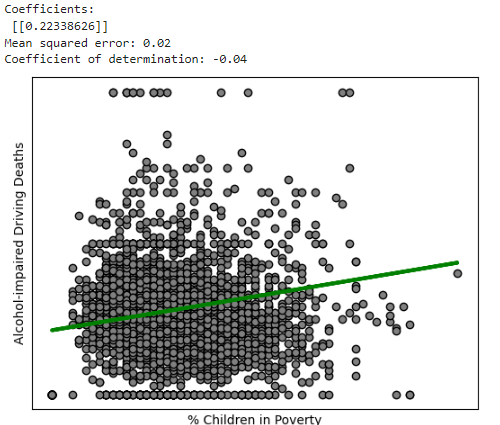
Chart, scatter chart

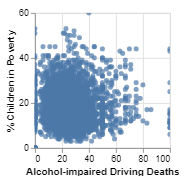
Description automatically generated



The figures above show the trend line for the % children in poverty & % frequent physical distress. This shows an increasing correlation between the two. This information led me to inspect further how each health behavior was correlated with the individual health behaviors and health outcomes. it's important to understand that poverty is a significant social determinant of health that can impact health outcomes and behaviors. People living in poverty are more likely to experience poor health outcomes and engage in unhealthy behaviors than those who are not living in poverty. The same can be said for the increased presence of physical distress. It can result in issues regarding mental and physical health, serving as a catalyst into unhealthy physical and mental health behaviors and outcomes.

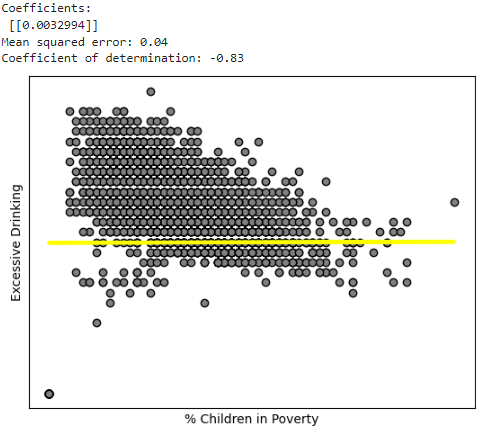
## Children in Poverty rate & Alcohol-impaired Driving Deaths

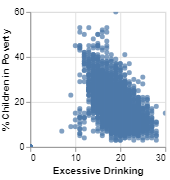




Above shows the prediction for the Child poverty rate as it correlates to the Alcohol-impaired driving death rate. The positive correlation shows an expected increase for both categories. These figures provide some evidence to our initial claim and show a need to push further into the research. Exploring the correlation between health behaviors and how those behaviors affect individual health outcomes, you can gain a deeper understanding of how specific behaviors and actions impact overall health. With the trend line showing the increase in Alcohol-impaired driving deaths, then I think it is fair to say that an increase in physical distress may result in an increase in Alcohol-impaired driving death seeing as which the childhood poverty rate is closely linked to overall physical distress, and both of those categories are linked to a positive trend in Alcohol-impaired driving deaths.

## Children in Poverty rate & Excessive drinking





The figures above show the trend line for the % children in poverty & the excessive drinking rate. This shows a neutral or potentially negative correlation between the two. This information comes a bit surprising based on the previous results that showed a positive trend between Alcohol-impaired driving deaths and the rate of children in poverty. This is most likely not as expected because this is the rate of children in poverty relating to excessive drinking. When in mind I was considering the rate of children in poverty and that relating to drinking excessively later on in life. The relationship between poverty and excessive drinking is likely complex, and may be influenced by a variety of factors. For example, children in poverty can be associated with increased levels of physical distress stress, which could contribute to higher rates of excessive drinking. However, poverty may also limit access to alcohol or reduce the ability to purchase it, which could have a mitigating effect on excessive drinking rates. And or this example, Children don’t have the ability to purchase alcohol so I think that serves to explain the negative correlation. Overall, these results highlight the need for further research to better understand the complex relationship between poverty and alcohol consumption, as well as the factors that contribute to these patterns over time.

## Children in Poverty rate & Alcohol-impaired Driving Deaths

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

The figures above show the trend line for the % children in poverty & the excessive drinking rate. This shows a neutral or potentially negative correlation between the two. This information comes a bit surprising based on the previous results that showed a positive trend between Alcohol-impaired driving deaths and the rate of children in poverty. This is most likely not as expected because this is the rate of children in poverty relating to excessive drinking. When in mind I was considering the rate of children in poverty and that relating to drinking excessively later on in life. The relationship between poverty and excessive drinking is likely complex, and may be influenced by a variety of factors. For example, children in poverty can be associated with increased levels of physical distress stress, which could contribute to higher rates of excessive drinking. However, poverty may also limit access to alcohol or reduce the ability to purchase it, which could have a mitigating effect on excessive drinking rates. And or this example, Children don’t have the ability to purchase alcohol so I think that serves to explain the negative correlation. Overall, these results highlight the need for further research to better understand the complex

## Children in Poverty rate & Quality of life

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

The figures above show the relationship between the rate of children in poverty and the effect it has on the quality of life. This correlation between the two shows the affective nature of the children in poverty. The consensus on the matter at hand is that poverty in general affects the quality of one’s life in a negative manner. So it’s surprising to see that the predictive analysis I ran using the data set shows that as the percentage of children in poverty increases so does the quality of life. This seems hard to believe and I would like to rule this out as inconclusive. Of course, there are other factors that come into play and when regarding the affects that poverty has on the quality of life so I would say that it’s important to remember that correlation doesn’t equal causation and that this finding is a good indicator of more research needing to be conducted on the factors that tie into the quality of life.

## Children in Poverty rate & Length of Life

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Again, The figures above show the relationship between the rate of children in poverty and the effect it has on the length of life. This correlation between the two shows the affective nature on the rate of children in poverty. The consensus on the matter at hand is that poverty in general affects the quality and length of one’s life in a negative manner. So it’s surprising to see that the predictive analysis ran using the data set shows that as the percentage of children in poverty increases so does the length of life. This seems hard to believe and I would like to rule this out as inconclusive. Again, there are other factors that come into play and when regarding the affects that poverty has on the length of life so I would say that it’s important to remember that correlation doesn’t equal causation and that this finding is a good indicator of more research needing to be conducted on the factors that tie into the length of life and the factors affecting it.

After running the predictions with the rate of frequent physical distress as the x-axis factor the results were identical in regards to the trends in prediction.

# DATA RESEARCH METHOD

The method to accrue results from most of the data could best be defined as clustering, like the machine learning technique. This is because the data was broken down into its different subsections and correlations and conclusion were drawn from the findings. The name I would give it is frame clustering as it was all achieved through breaking down the data in to correlative data frames and using the built in functions and tools to draw conclusion on the relationships.

# EVALUATIONS

Upon evaluating the data and the results, I would say that the project is much like a puzzle and you’re working to find how each piece of the data connects to one another. There were definitely relationships and findings in the data that I didn’t expect. I think this project improved my analytical skills tremendously and there is so much more room to explore the datasets for connections. I found Data Science to be a very interesting field and I intend to learn more about the techniques and applications of data science.

# Conclusions and Future Work

**Q1:**

There appears to be no distinct correlation between broadband access and suicide rate. The states with the four highest suicide rates (Wyoming, New Mexico, New York, and California), all have different rates of broadband access, and on top of that, the broadband access in every state is high to the point where it doesn't seem like any small dip of a percent has a major effect.

**Q2:**

It appears that there is some correlation between single parent households and teen birth rate. Specifically, looking at states with higher teen birth rates (Mississippi, Oklahoma, etc.) They tend to have high/midrange single parent household rates. When it comes to single parent households and arrest rates, the results are a bit inconclusive due to a lack of stats provided by the dataset, but it appears there is some correlation there from the data that is there. States with lower juvenile arrest rates tend to be in the low/mid-range for single parent households, whereas states with high juvenile arrest rates tend to be in the mid to high range for single parent households.

**Q3:**

There isn't any direct correlation between the percentage of insufficient sleep and the high school graduation rate because the areas that reported the highest level of sleep inefficiency, also have the highest graduation rate in the country, but the graduation rate is high for majority of the country. There does appear to be some correlation between sleep inefficiency and the percentage of disconnected youth. This can mainly be shown in counties in the states of New Mexico, Nevada, Oklahoma, Texas, Florida, and Michigan.

**Q4:**

There didn't appear to be a correlation between access to healthy food and either of the two categories. But there did appear to be a direct correlation between the percentage of physically inactive people and the percentage of being frequently distressed to the point where the maps nearly look identical. This is most prominently shown in the states of Texas, Mississippi, Oklahoma, Georgia, and West Virginia.

**Q5:**

Like question one, where it might be assumed that lower homeowner rate would lead to higher suicide rates, based on the bar graphs it appears that would be false. Looking at the lowest homeowner percent (DoC), they also have the lowest rate of suicide. And even places with much higher homeowner percentages like West Virginia or Louisiana still have a high suicide rate, almost twice so compared to the District of Columbia. The states with the highest suicide rates (Wyoming, New Mexico, New York, and California) all also have decent homeownership precents.

# CONCLUSION

The first model created aimed to explore the association between the rate of children living in poverty and their frequency of physical distress. Through the use of a heat map, the initial investigation revealed that the two variables were highly correlated. In other words, areas with a higher percentage of children living in poverty tended to have higher rates of physical distress among children as well. Upon further analysis, the predictive model confirmed these results, indicating that a higher prevalence of child poverty was a strong predictor of increased physical distress. These findings are significant for several reasons. Firstly, higher rates of physical distress are associated with a range of negative health outcomes, including a greater likelihood of chronic conditions, reduced quality of life, and decreased life expectancy. Additionally, physical distress may also lead to the development of negative health behaviors, such as smoking, drinking, or drug abuse, which further exacerbate negative health outcomes. Furthermore, children living in poverty may experience a range of adverse health outcomes beyond physical distress. For example, they may be at an increased risk of poor mental health, malnutrition, or developmental delays, all of which can have long-lasting effects on their overall health and wellbeing. Given these consequences, it is imperative that we prioritize efforts to reduce child poverty and alleviate its associated health burdens. Effective strategies may include increasing access to affordable housing, improving education and job training opportunities, and implementing policies that ensure all families have access to quality healthcare. By taking these steps, we can promote better health outcomes for children, reduce the incidence of physical distress and its negative consequences, and create a healthier and more equitable society.

# FUTURE WORK

In the future I plan to continue working on this project and using future data sets to conducts personal project and find results. I could have marginalized the results and seen which race is most affected by these socioeconomic factors which I plan to do in my free time. This could present more insight into which minorities are most affected. I plan to enter competitions on Kaggle as well to buff my data organization skills and I hope to get more experience using data.

# Bibliography

Rankings, C. H. (2022). *County Health Rankings & Roadmaps*. Retrieved from County Health Rankings & Roadmaps: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation

*scikit learn*. (n.d.). Retrieved from scikit learn: https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LinearRegression.html