# Association Between Poverty and Obesity in the United States

## Abstract

A recent survey of U.S. adults showed that one of the most common New Year's resolutions for 2018 was to lose weight or get into shape. According to a recent food and health survey, 52% of Americans polled believe it is easier to do their taxes than to figure out how to eat healthfully. Knowing how complicated our tax codes are, this means that eating healthy can really, really be tricky for many people. As a country in a battle with obesity, healthy eating is not easy. Over a third of the American population is classified as obese; this means that almost 41 million women and more than 37 million men aged 20 and older were obese in 2010. However, obesity statistics do not encompass the entire picture. When we combine statistics on Americans who are obese or overweight, we find that more than 2 in 3 adults are overweight or obese.

According to the U.S. Centers for Disease Control and Prevention, over one-third of U.S. adults are obese. CDC defines obesity as a body mass index equal to or greater than 30. In simpler terms, a person 5 feet 4 inches tall is obese if she or he weighs 174 pounds or more; a person 5 feet 9 inches is obese if she or he weighs 203 pounds or more.

Obesity is related to some of the leading causes of death, including heart disease, some cancers, stroke, and type 2 diabetes. While obesity levels have been rising for all socioeconomic groups, some groups are more affected than others. Recent research highlights the complexity and variation in how socioeconomic status and obesity are related.

A study published in Social Science and Medicine used data for 67 countries representing all the regions of the world to examine how economic development, socioeconomic status, and obesity were related. The researchers used self-reported height and weight to calculate body mass index (weight relative to height), and looked at the relationship between obesity, gross national product, and SES such as education, occupation, and income. They found that obesity rose with a nation's economic

development, but also that socioeconomic status as it related to obesity changed. In lower-income countries, people with higher SES were more likely to be obese. Conversely, in high-income countries, those with higher SES were less likely to be obese.

## Introduction

There are so many theories about overweight and obesity in the United States. I travel internationally for both work and pleasure. I notice that there are three main reasons why the obesity rate is so high in the United States compare to other countries. The number one reason must be engineered junk foods. Heavily processed foods are often little more than refined ingredients mixed with additives. These products are designed to be cheap, last long on the shelf and taste so incredibly good that they are hard to resist. By making foods as tasty as possible, food manufacturers are trying to increase sales. But they also promote overeating. The second reason for high obesity rate in the United States is food addiction. Many sugar-sweetened, high-fat junk foods stimulate the reward centers in your brain. In fact, these foods are often compared to commonly abused drugs like alcohol, cocaine, nicotine, and cannabis. Junk foods can cause addiction in susceptible individuals. These people lose control over their eating behavior, like people struggling with alcohol addiction losing control over their drinking behavior. Addiction is a complex issue that can be very difficult to overcome. When you become addicted to something, you lose your freedom of choice and the biochemistry in your brain starts calling the shots for you. Finally, we can refer to food availability as the last factor that contributes to the obesity problem in the United States. Food availability dramatically influences people's waistline, which has increased massively in the past few centuries. Food, especially junk food, is everywhere now. Shops display tempting foods where they are most likely to gain your attention. Another problem is that junk food is often cheaper than healthy, whole foods, especially in America. Some people, especially in poorer neighborhoods, do not even have the option of purchasing real foods, like fresh fruit and

vegetables. Convenience stores in these areas only sell sodas, candy and processed, packaged. Most processed foods today do not resemble whole foods at all. These are highly engineered products, designed to get people hooked.

## Methods

### Data sources

This project is very special. I have been working on it for about six months now, and I am very excited that I will professionally complete it. I have used government statistics because national governments are often the only institutions with the resources to collect comprehensive social statistics, and thus publish many social statistics available. There are two datasets that I will examine. The primary data has not been generated by surveys, interviews, and experiments. It is a normal dataset that was generated from Data.gov, and it is designed for understanding and solving the research problem at hand. The secondary data truly follows the definition of a real secondary data. It is generated by census.gov. It will serve as supporting data for the project.

- Data source: https://catalog.data.gov/dataset/national-obesity-by-state
- Data source: https://www.census.gov/topics/income-poverty/poverty/data/tables.html

## Association Between Poverty and Obesity in the United States

| State | Percent in Obesity | Percent in poverty | Hispanic | White | Black | Asian | American Indian |
|---|---|---|---|---|---|---|---|
| AL | 35.6 | 16.3 | 4.10% | 65.50% | 26.70% | 1.30% | 0.50% |
| AK | 29.8 | 10.4 | 7.00% | 60.60% | 2.90% | 6.60% | 14.20% |
| AZ | 28.4 | 18.2 | 31.40% | 54.70% | 4.10% | 3.20% | 3.90% |
| AR | 34.3 | 21.2 | 7.40% | 72.30% | 15.20% | 1.60% | 0.60% |
| CA | 24.2 | 17 | 39.10% | 37.00% | 5.50% | 14.40% | 0.40% |
| CO | 20.2 | 10.9 | 21.50% | 68.20% | 3.90% | 3.10% | 0.60% |
| CT | 25.3 | 8.6 | 16.10% | 66.70% | 9.90% | 4.50% | 0.20% |
| DE | 29.7 | 10.5 | 9.30% | 62.20% | 21.50% | 4.00% | 0.20% |
| DC | 22.1 | 30.7 | 11.00% | 36.50% | 45.30% | 4.00% | 0.20% |
| FL | 26.8 | 12.9 | 25.60% | 53.80% | 15.40% | 2.80% | 0.20% |
| GA | 30.7 | 18.3 | 9.60% | 52.60% | 31.10% | 3.90% | 0.20% |
| HI | 22.7 | 9 | 10.50% | 21.80% | 1.60% | 37.30% | 0.10% |
| ID | 28.6 | 11.8 | 12.40% | 82.00% | 0.60% | 1.30% | 1.10% |
| IL | 30.8 | 14 | 17.20% | 61.20% | 14.00% | 5.40% | 0.10% |
| IN | 31.3 | 11.3 | 6.90% | 79.20% | 9.20% | 2.20% | 0.10% |
| IA | 32.1 | 13 | 5.90% | 85.90% | 3.30% | 2.60% | 0.20% |
| KS | 34.2 | 16.4 | 11.90% | 75.90% | 5.50% | 2.90% | 0.60% |
| KY | 34.6 | 22.6 | 3.50% | 84.60% | 8.00% | 1.40% | 0.20% |
| LA | 36.2 | 23 | 5.20% | 58.50% | 32.10% | 1.80% | 0.50% |
| ME | 30.8 | 10.9 | 1.60% | 93.40% | 1.20% | 1.10% | 0.60% |
| MD | 28.9 | 9.1 | 10.10% | 50.70% | 29.40% | 6.40% | 0.20% |
| MA | 24.3 | 12.2 | 11.80% | 71.50% | 7.00% | 6.60% | 0.10% |
| MI | 31.2 | 17.8 | 5.10% | 75.00% | 13.60% | 3.10% | 0.50% |
| MN | 26.1 | 9.9 | 5.30% | 79.90% | 6.40% | 4.90% | 1.00% |
| MS | 35.6 | 29.6 | 2.90% | 56.60% | 37.90% | 0.90% | 0.40% |
| MO | 32.4 | 15.8 | 4.20% | 79.40% | 11.40% | 2.00% | 0.30% |
| MT | 23.6 | 15.2 | 3.70% | 86.30% | 0.40% | 0.70% | 5.90% |
| NE | 31.4 | 12.3 | 10.90% | 79.00% | 4.50% | 2.40% | 0.70% |
| NV | 26.7 | 11 | 28.80% | 48.80% | 8.90% | 8.30% | 0.90% |
| NH | 26.3 | 4.9 | 3.80% | 90.30% | 1.30% | 2.70% | 0.10% |
| NJ | 25.6 | 11.1 | 20.40% | 54.80% | 12.80% | 9.80% | 0.10% |
| NM | 28.8 | 21.6 | 48.80% | 37.40% | 1.80% | 1.30% | 8.80% |
| NY | 25 | 18.9 | 19.20% | 55.10% | 14.30% | 8.70% | 0.20% |
| NC | 30.1 | 17.9 | 9.40% | 63.00% | 21.20% | 2.90% | 1.10% |
| ND | 31 | 11.4 | 3.50% | 84.40% | 3.00% | 1.70% | 5.40% |
| OH | 29.8 | 16.1 | 3.70% | 78.90% | 12.20% | 2.20% | 0.20% |
| OK | 33.9 | 19.2 | 10.60% | 65.60% | 7.20% | 2.10% | 7.30% |
| OR | 30.1 | 15.2 | 13.10% | 75.60% | 1.80% | 4.30% | 0.90% |
| PA | 30 | 15.6 | 7.30% | 76.40% | 10.70% | 3.50% | 0.10% |
| RI | 26 | 15.3 | 15.40% | 72.10% | 5.40% | 3.60% | 0.30% |
| SC | 31.7 | 14.1 | 5.70% | 63.60% | 26.80% | 1.50% | 0.20% |
| SD | 30.4 | 13.3 | 3.60% | 82.30% | 1.90% | 1.20% | 8.60% |
| TN | 33.8 | 19.1 | 5.40% | 73.90% | 16.60% | 1.80% | 0.20% |
| TX | 32.4 | 20.4 | 39.40% | 41.90% | 11.80% | 4.80% | 0.30% |
| UT | 24.5 | 11.5 | 14.00% | 78.30% | 1.20% | 2.40% | 1.00% |
| VT | 25.1 | 8.1 | 1.90% | 92.80% | 1.20% | 1.80% | 0.30% |
| VA | 29.2 | 11 | 9.30% | 61.70% | 18.80% | 6.40% | 0.20% |
| WA | 26.4 | 10.4 | 12.70% | 68.60% | 3.50% | 8.50% | 1.00% |
| WV | 35.6 | 20.2 | 1.30% | 92.00% | 3.90% | 0.80% | 0.10% |
| WI | 30.7 | 14.7 | 6.90% | 81.20% | 6.30% | 2.70% | 0.80% |
| WY | 29 | 12 | 10.00% | 84.00% | 0.90% | 0.80% | 2.10% |

### Statistical methods

I have decided to examine the association between obesity independent variables and state percentage of poverty prevalence by using ordinary least squares (OLS). OLS is a variation of linear regression, a statistical method that examines associations between multiple independent variables and a single dependent variable; once the assumptions are satisfied, the regression output indicates the strength of the association between the dependent variable and each of the independent variables. These

assumptions, include linear parameters, random sampling, no multicollinearity, no autocorrelation, a conditional mean of zero, and normally distributed error terms; all of them were satisfied, meaning that our OLS models are efficient and represent a linear unbiased estimator of variable coefficients.

I also believe that simple linear regression may be a great way to examine my single input. Because simple linear regression requires statistical properties from the data such as means, standard deviations, correlations, and covariance, all the data must be available to traverse and calculate statistics.

## Visualization methods

I will use R and Power BI for this project. I am not very comfortable using R, but I am sure that its applications are very well aligned with this type of projects. I will use ggplot which gives me a coherent way to produce visualizations by expressing relationships between the attributes of data and their graphical representation. I will use R Markdown to allow me to create documents that serve as a neat record of your analysis. And finally, I may use Dplyr for data manipulation if I need it.

I will also use Power BI for the following reasons.

- Power BI offers a wide range of custom visualizations. That means visualizations made by developers for a specific use. Custom visuals are also available on Microsoft marketplace.
- In Power BI, I also have the option to upload and view my data in Excel. I will be able to select, filter, and slice data in a Power BI report or dashboard and put it on Excel. I will be able to open Excel and view the same data in tabular form in an Excel spreadsheet.
- Another reason why I will use Power BI is because the data visualization tool is very *attractive, intuitive, and interactive.* It is very easy to create and understand data through visualizations in Power BI.

## Explain the Datasets

As I mention above, there are two datasets that will be used and analyzed in this project.

## Association Between Poverty and Obesity in the United States

1. The first data is called “National Obesity by State”. This dataset will be analyzed to determine the states that have the greater obesity rate in the United States.
2. The second data is called “Poverty by State”. This dataset will be used almost the same way as National Obesity by State. The goal of this dataset is to use it to prove that poverty plays a big role in obesity in the United States.

### Data Analysis and Research Questions

- Are poverty and obesity associated?

  Poverty rates and obesity were reviewed across all 51 states in the U.S. In contrast to international trends, people in America who live in the most poverty-dense counties are those most prone to obesity. States with poverty rates of greater than 35% have obesity rates 145% greater than wealthy counties.

- How is poverty linked to obesity?

  I will prove that individuals who live in impoverished regions have poor access to fresh food. Poverty-dense areas are oftentimes called “food deserts,” implying diminished access to fresh food.

  I will also prove the evidence of the association between sedentariness, poor health, obesity, diabetes, other metabolic diseases, and premature death. Sedentary individuals move 2 hours per day less than active individuals and expend less energy, and they are thereby prone to obesity, chronic metabolic disease, and cardiovascular death.

- What can be done to solve the obesity problem?

  I will create focus on predictive analysis to pinpoint a diverse set of physical activities and nutritional programs in neighborhoods and educational institutions may be the best way to reduce the risk of obesity among poor youngsters. I will also some such that need reliable

funding from governments and charities to boost and leverage valuable programs in neighborhood centers and schools serving disadvantaged Americans.

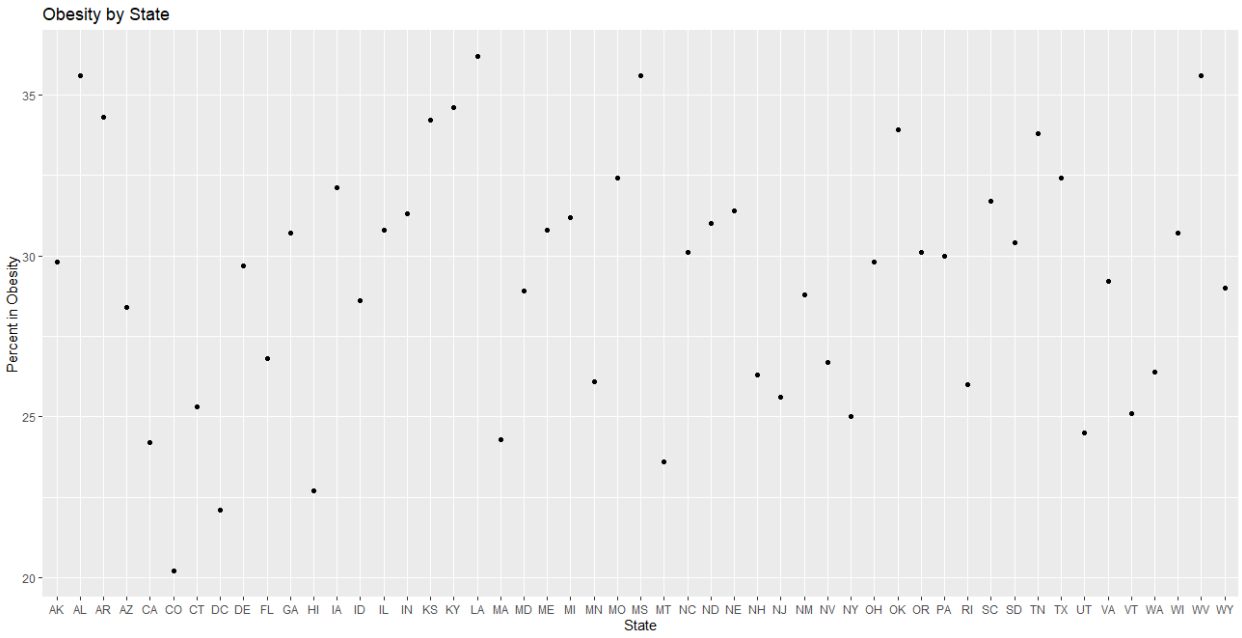
## Results

```
# Import libraries
library(lattice)
library(ggplot2)
library(ppcor)

# Import the week-7-housing
project1_data<- read.delim(file.choose(), header=T, sep=",")
# Attach the data
attach(project1_data)

project1_data

summary(project1_data)
summary(Percent.in.Obesity)
summary(Percent.in.poverty)

ggplot(project1_data, aes(x = State, y = Percent.in.Obesity)) +
  geom_point() +
  labs(x = 'State', y = 'Percent in Obesity', title = 'Obesity by State')

ggplot(project1_data, aes(x = State, y = Percent.in.poverty)) +
  geom_point() +
  labs(x = 'State', y = 'Percent in Poverty', title = 'Poverty by State')
```

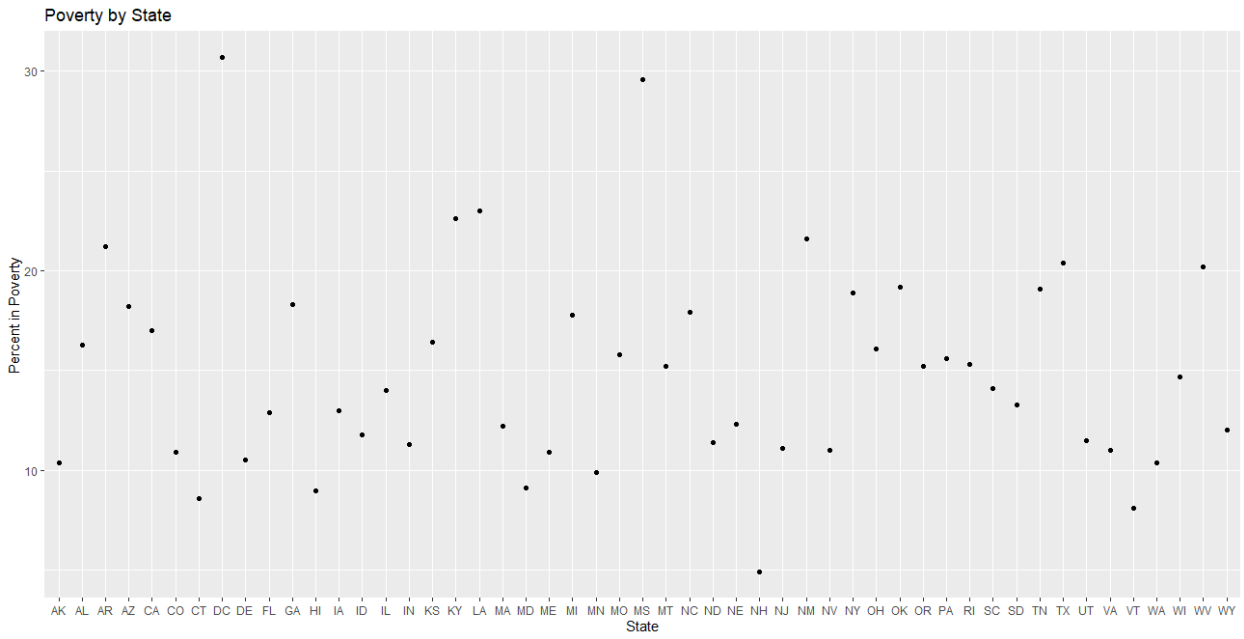
```
ggplot(project1_data, aes(x = State, fill = Hispanic, White, Black, Asian, American.Indian)) +
 geom_bar() +
 labs(x = 'State', y = 'Percent in Population', title = 'Population by State')

#simple regression
cov(Percent.in.Obesity, Percent.in.poverty, method = "pearson")

cor.test(Percent.in.Obesity, Percent.in.poverty, method = "pearson", alt="greater",
conf.level=0.99)

t.test(Percent.in.Obesity, Percent.in.poverty, mu=0, alt="two.sided", paired=T, conf.level=0.99)

cor.test(x = Percent.in.Obesity, y = Percent.in.poverty)

pcor.test(x = Percent.in.Obesity, y = Percent.in.poverty, z = White)

var(Percent.in.Obesity)
var(Percent.in.poverty)

var(Percent.in.Obesity)/var(Percent.in.poverty)
var.test(Percent.in.Obesity,Percent.in.poverty)

project1_data1 <- lm(Percent.in.Obesity ~ Percent.in.poverty)
summary(project1_data1)

N<- length(Percent.in.Obesity)
N
a<- mean(Percent.in.Obesity)
```

```
a
b<- mean(Percent.in.poverty)
b
c<- length(Percent.in.poverty)
c

confint(project1_data1, conf.level=0.95)
plot(project1_data1)

# calculate the standardized residuals
Residuals<- residuals(project1_data1)
predict(project1_data1)
plot(Percent.in.Obesity, Residuals)

# The sum of large residuals.
sum(Residuals)
residuals(project1_data1)^2
sum(residuals(project1_data1)^2)

#Simple linear Regression
plot(Percent.in.Obesity ~ White, data = project1_data)
plot(Percent.in.poverty ~ White, data = project1_data)
plot(Percent.in.poverty ~ Hispanic, data = project1_data)
plot(Percent.in.Obesity ~ Hispanic, data = project1_data)
plot(Percent.in.poverty ~ Black, data = project1_data)
plot(Percent.in.Obesity ~ Black, data = project1_data)
plot(Percent.in.Obesity ~ Percent.in.poverty, data = project1_data)

# Ordinary Least Squares Regression
```

```
olsreg<-lm(Percent.in.Obesity~Hispanic+Black+White)
summary(olsreg)

plot(olsreg)

olsregi<-lm(Percent.in.poverty~Percent.in.poverty)
summary(olsregi)

plot(olsregi)
```

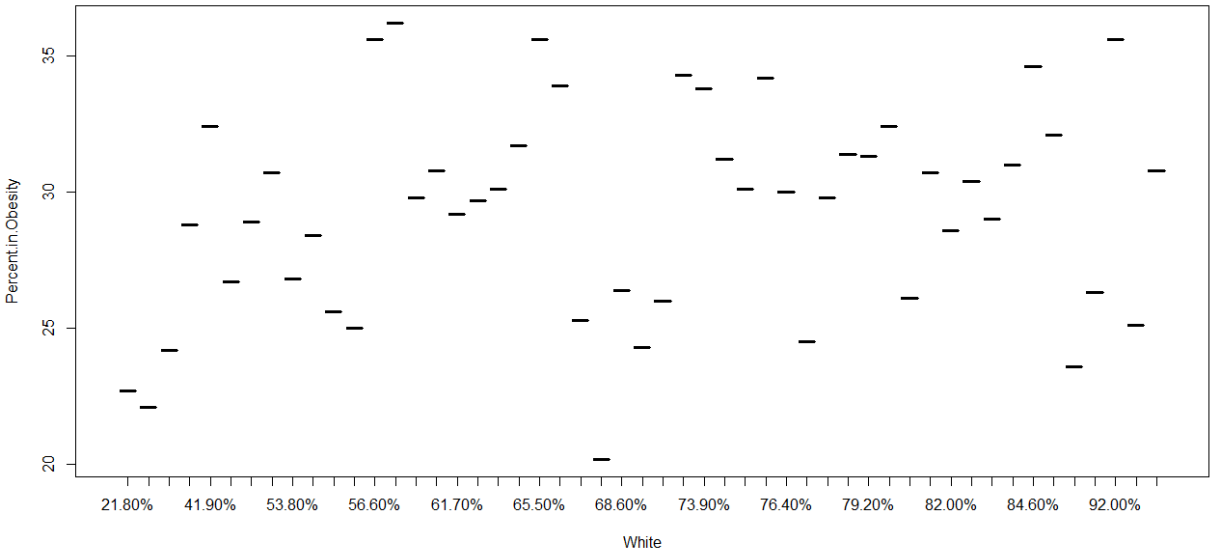


## Association Between Poverty and Obesity in the United States

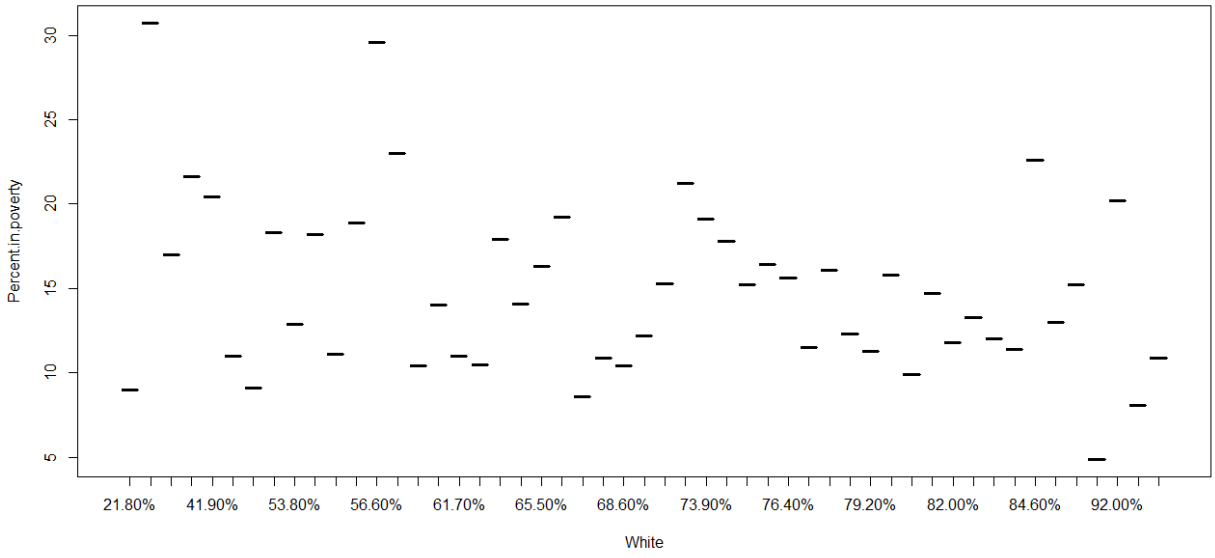


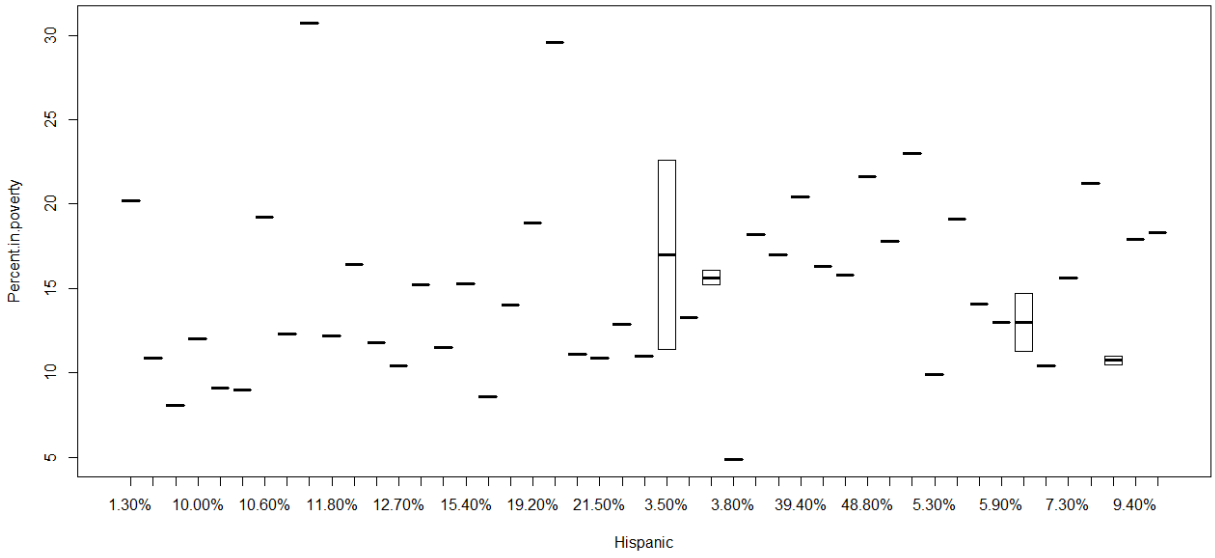

### Simple Regression and Ordinary Least Square Regression

As I mention in the statistical method section, I am using simple linear regression and ordinary least squares for my analysis. Well, I can clearly conclude that ordinary least squares regression was not helpful at all for the statistical analysis. I was kind of was happy with the result although I was expecting the correlation between obesity and poverty to be higher. As you can notice, the correlation is 0.4057821.



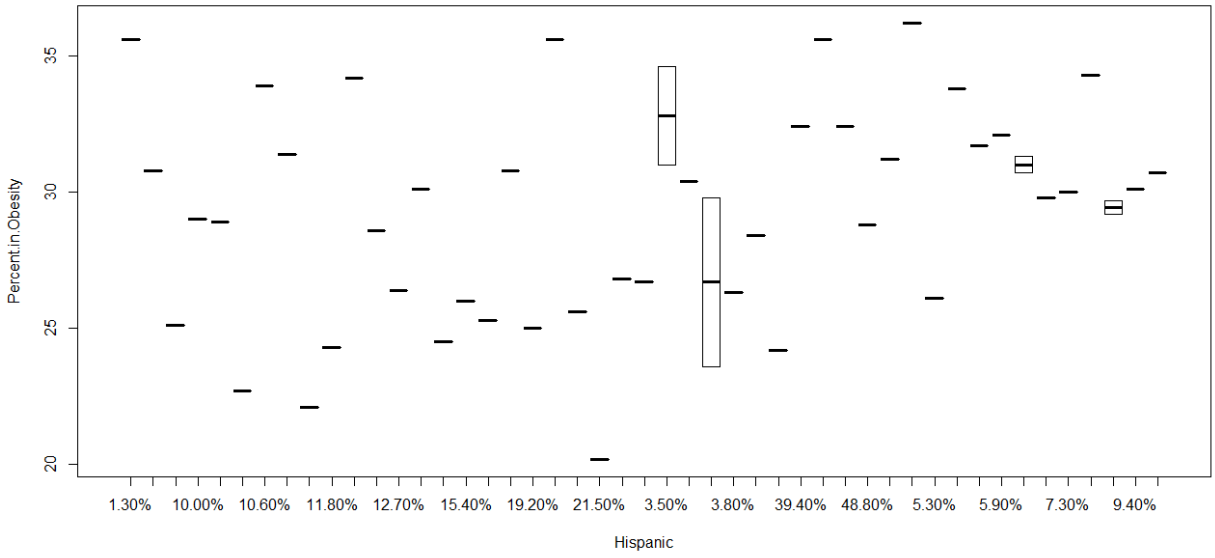
# Association Between Poverty and Obesity in the United States

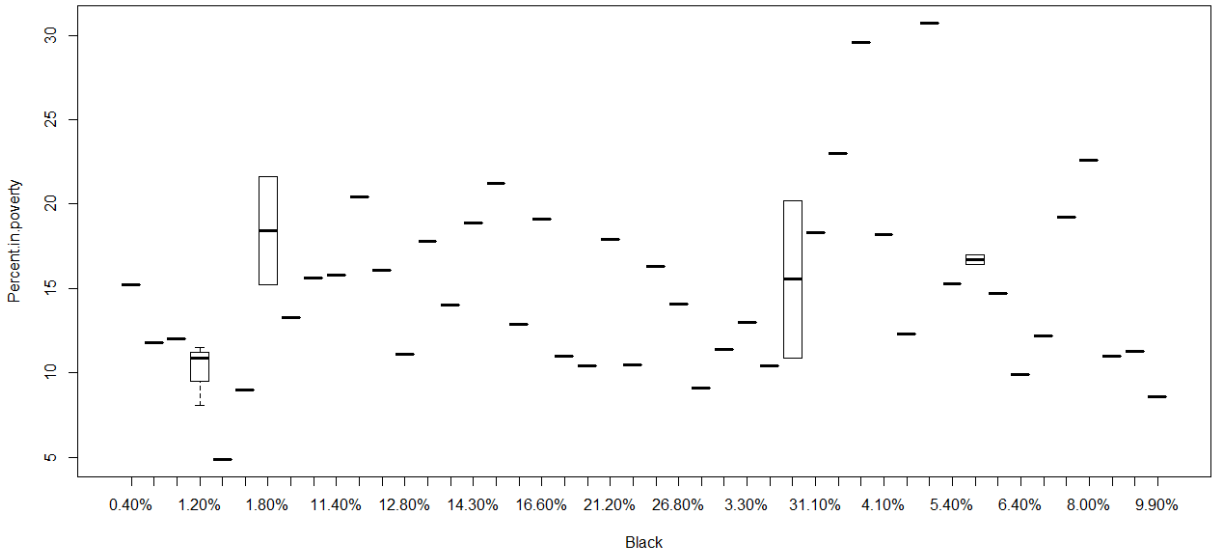


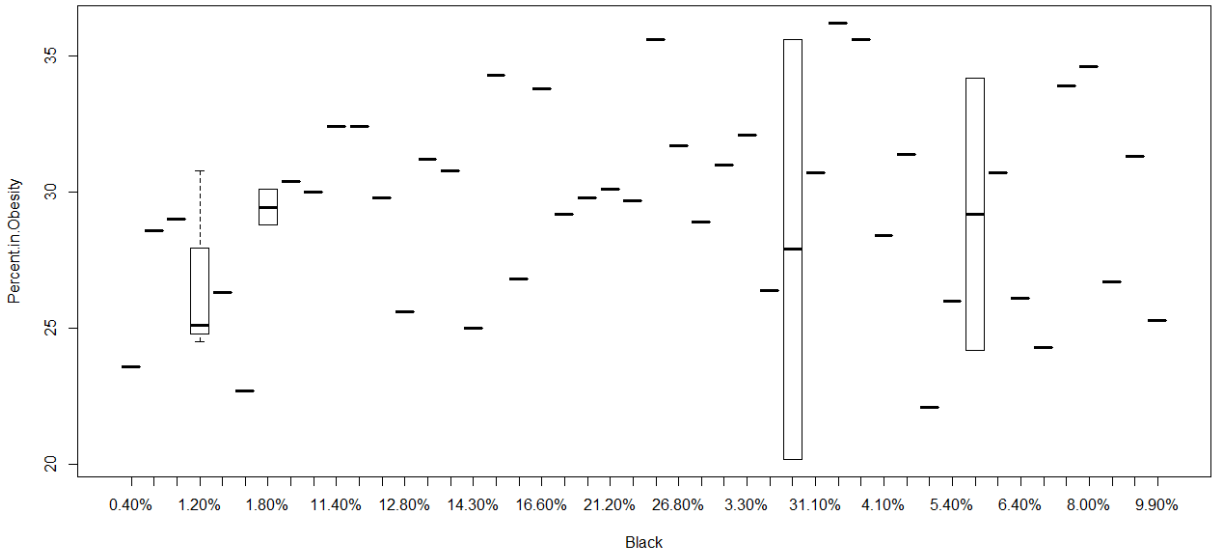

Luma, Alberto | BELLEVUE UNIVERSITY

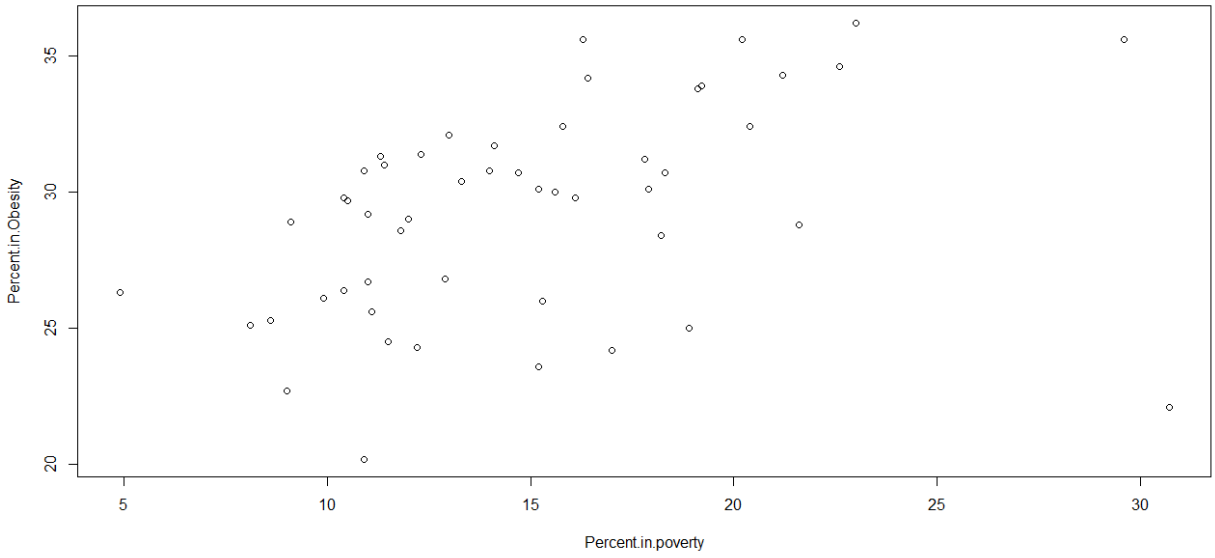
# Association Between Poverty and Obesity in the United States



## Association Between Poverty and Obesity in the United States



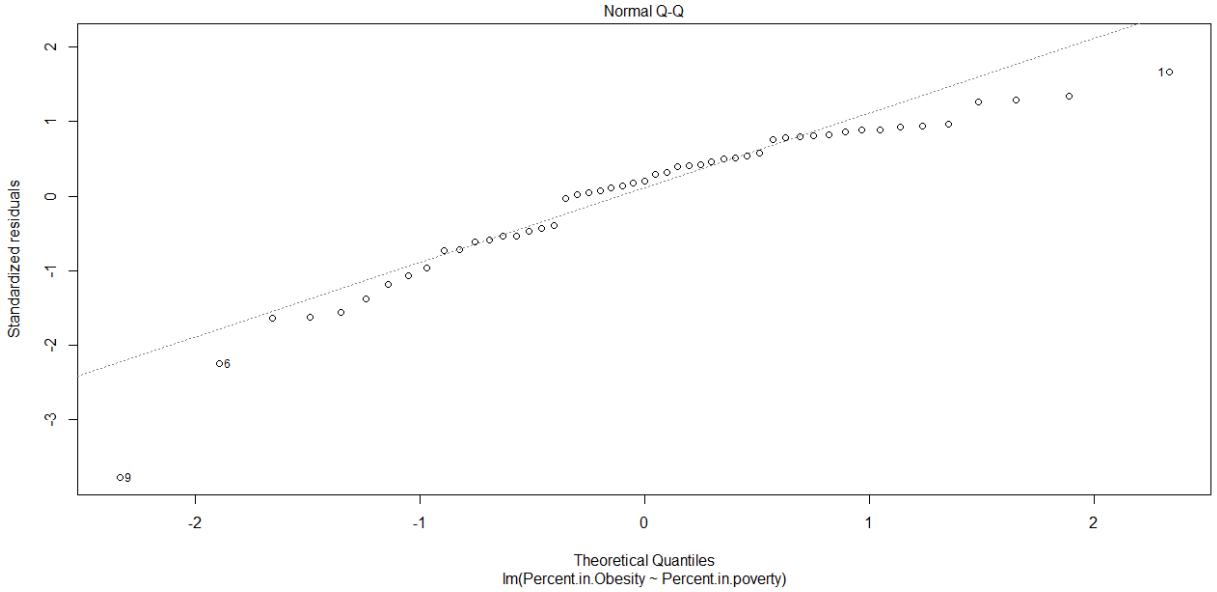



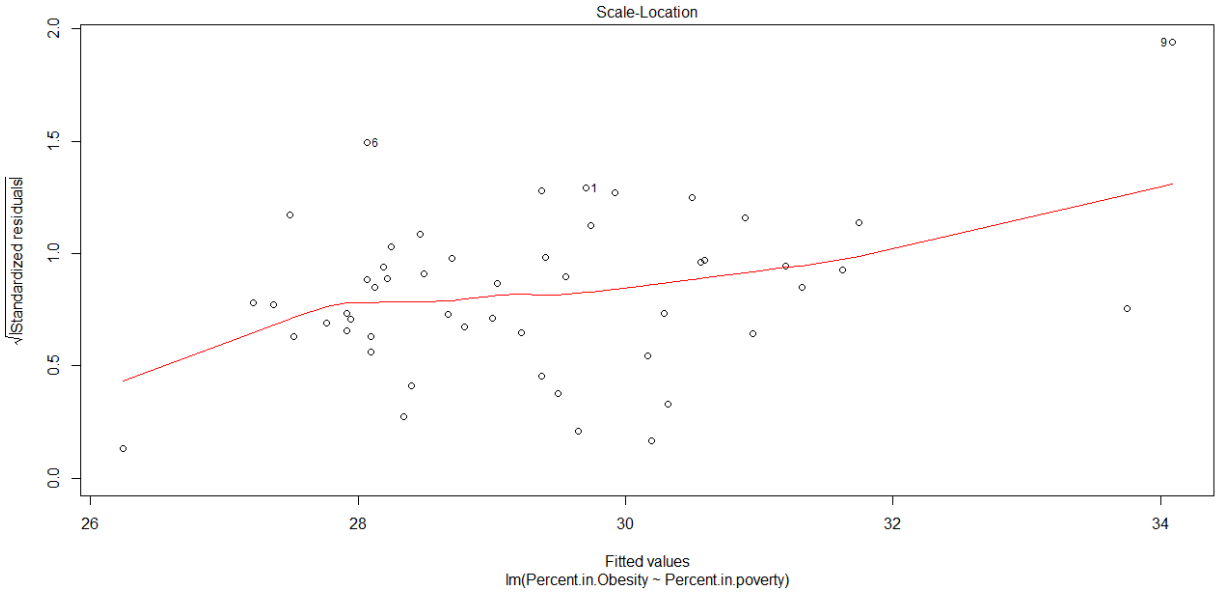

Pearson's product-moment correlation

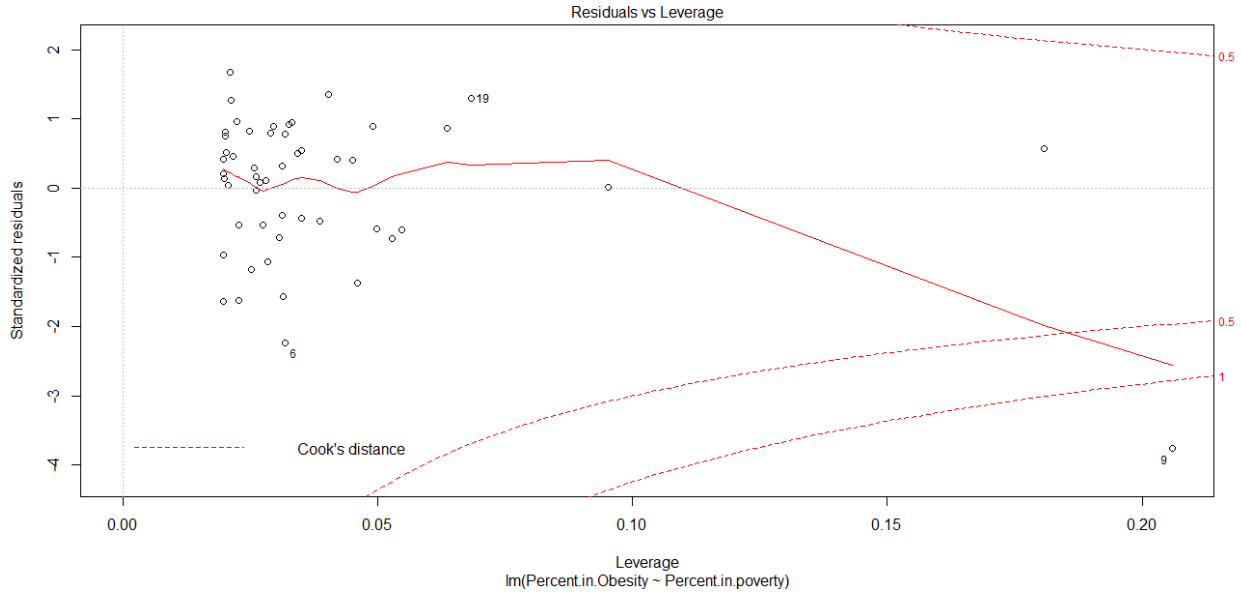
```
data:  Percent.in.Obesity and Percent.in.poverty
t = 3.1078, df = 49, p-value = 0.001567
alternative hypothesis: true correlation is greater than 0
99 percent confidence interval:
 0.09448939 1.00000000
sample estimates:
      cor
0.4057821
```

Paired t-test

```
data:  Percent.in.Obesity and Percent.in.poverty
t = 20.328, df = 50, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 12.46397 16.24584
sample estimates:
mean of the differences
                14.3549

var(Percent.in.Obesity)
[1] 14.97816

var(Percent.in.poverty)
[1] 26.67523

                        F test to compare two variances

data:  Percent.in.Obesity and Percent.in.poverty
F = 0.5615, num df = 50, denom df = 50, p-value = 0.04374
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3204999 0.9837232
sample estimates:
ratio of variances
         0.5615008

mean(Percent.in.Obesity) 29.29412
mean(Percent.in.poverty) 14.93922

                           2.5 %     97.5 %
(Intercept)        21.6470293 27.8561859
Percent.in.poverty  0.1074527  0.5006797
```

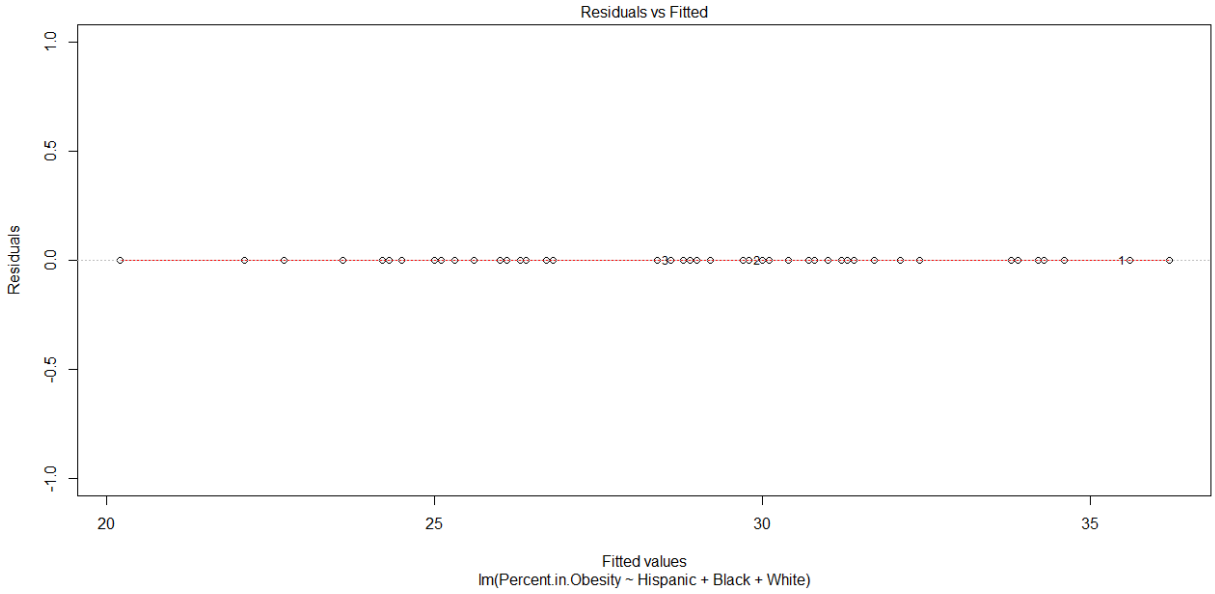


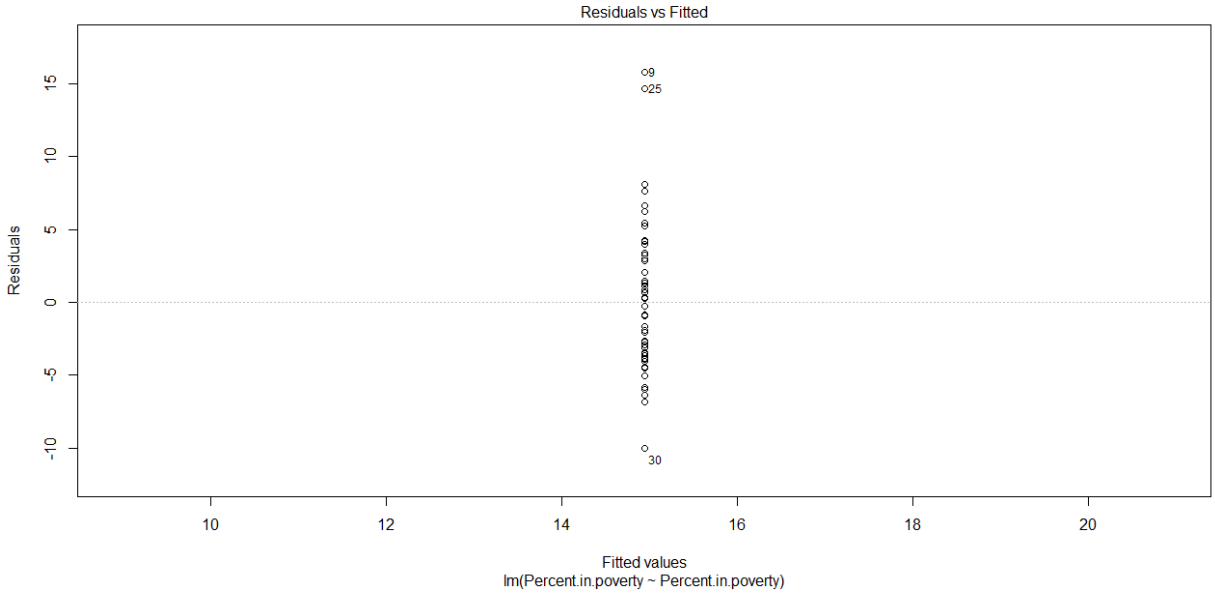



```
summary(olsregi)

Call:
lm(formula = Percent.in.poverty ~ Percent.in.poverty)

Residuals:
     Min       1Q   Median       3Q      Max
-10.0392  -3.8892  -0.8392   3.1108  15.7608

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.9392     0.7232   20.66   <2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.165 on 50 degrees of freedom
```




## Acknowledgments

There are so many articles, data, and organizations that I owe great credit and great deals of respects to. Like I mentioned above, I have started the projects with the simple idea to compare obesity rate across the US. After reading several articles on https://www.healthcare-administration-degree.net/poverty-

obesity/, I realized that I needed to add more concepts into my research to make my projects more appealing and concrete. These articles have really helped me to understand some of the more important ways to structure my analysis. I was able to collect more data on poverty by state.

Finally, I must thank my family especially my wife to allow me to skip so many family activities to focus on working in this project.

## Author Information



| | |
|---|---|
| Corresponding Author | Alberto Luma, 5802 Coquyt Drive, Mount Dora, Fl, 32757.<br>Email: lumaalberto@gmail.com |
| Author Affiliations | Bellevue University<br>DSC680 – Applied Data Science |



## References

1. https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity#estimation
2. https://catalog.data.gov/dataset/national-obesity-by-state
3. http://cs.carleton.edu/cs_comps/0910/netflixprize/final_results/knn/index.html
4. https://www.health.harvard.edu/staying-healthy/obesity-in-america-whats-driving-the-epidemic
5. https://www.sciencedirect.com/topics/computer-science/decision-trees
6. Food Service Guidelines Federal Workgroup. Food Service Guidelines for Federal Facilities. Washington, DC: U.S. Department of Health and Human Services, 2017. https://www.cdc.gov/obesity/downloads/ guidelines_for_federal_concessions_and_ vending_operations.pdf (accessed July 24, 2019).

7. "Operation Live Well." Health.mil, Military Health System. https://health.mil/Military-Health-Topics/Operation-Live-Well (accessed July 24, 2019).
8. American Academy of Pediatrics. "Promoting Food Security for All Children." Pediatrics, 135(5): e1431–e1438, 2015. https://pediatrics.aappublications.org/ content/136/5/e1431 (accessed July 24, 2019).
9. Reference: "The relationship between childhood obesity, low socioeconomic status, and race/ethnicity: Lessons from Massachusetts," Childhood Obesity, December 2015, Vol. 11, No. 6.
10. Additional authors: Robert Rogers; Taylor F. Eagle; Anne Sheetz, M.P.H.; Alan Woodward, M.D.; Robert Leibowitz, Ph.D.; MinKyoung Song, Ph.D.; Rachel Sylvester; Nicole Corriveau; Eva Kline-Rogers; Qingmei Jiang and Elizabeth Jackson, M.D., M.P.H.
11. https://www.governing.com/gov-data/census/state-minority-population-data-estimates.html

## Tables

| summary(Percent.in.Obesity) | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|---|
| | 20.20 | 26.20 | 29.80 | 29.29 | 31.55 | 36.20 |
| summary(Percent.in.Poverty) | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| | 4.9 | 11.05 | 14.10 | 14.94 | 18.05 | 30.70 |