|  |
| --- |
| **Dalhousie University** |
| **STAT 3340** |
| **Regression of Obesity Among Adults, by Income and Education** |
|  |
| Group 10  Yutong Liu B00728064  Yifan Chen B00745748  Date: December 5, 2020 |
|  |

## Abstract

This project is mainly through to the United States in 2011-2016 adults are overweight and obesity of sampling data is analyzed, American adult obesity epidemic characteristics, and related influence factors of overweight and obesity to establish regression model, find out the influence factors of obesity, for the future to carry out targeted intervention and prevention measures to provide some basis.

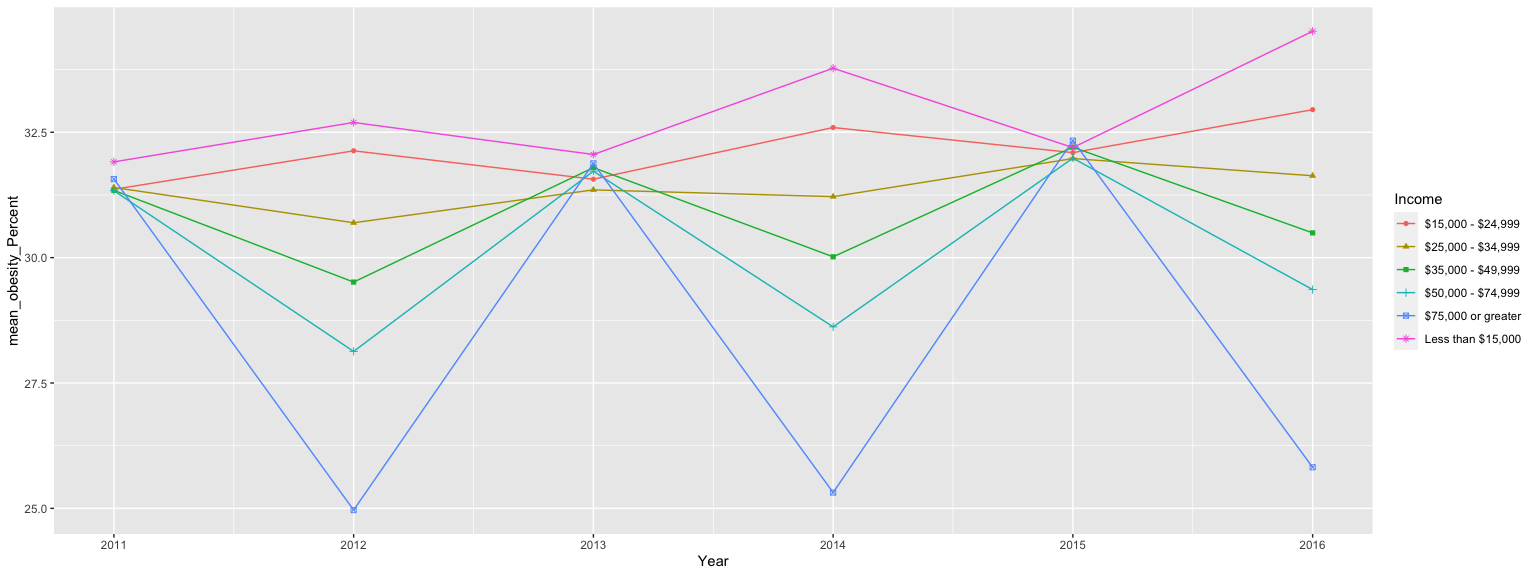
## Introduction

Obesity is a common, obvious and complex metabolic disorder syndrome. It refers to the disturbance of energy balance caused by long-term and chronic energy intake exceeding energy consumption, which leads to excessive energy storage in the form of fat, which is manifested as excessive accumulation and abnormal distribution of fat. It seriously affects the normal physiological function of the body, and has reached the level of damage to health (Lockwood, 2010). In 2000, the World Health Organization ranked obesity as one of the top three health killers, along with smoking and AIDS. In recent years, with the rapid development of science and technology and economy, great changes have taken place in people’s lifestyle and dietary structure, which makes the prevalence of overweight and obesity increase rapidly in both developed and developing countries all over the world, and the trend is increasing year by year. According to WHO data, more than 1.6 billion people were overweight or obese worldwide in 2005, while the United States had the highest prevalence of obesity, with an obesity rate of 32.0% (Centers for Disease Control and Prevention, n.d.). The American Health and Nutrition Examination Survey, which analyzed the data from five cross-sectional adult health surveys conducted between 1960 and 2000, found that over the 40 years, the overweight rate of adult residents in the United States did not increase significantly, but remained at a relatively high level, while the obesity rate showed an increasing trend. The obesity rate was 30.4% in 2000, an increase of 108.2% compared with that in 1960 (Centers for Disease Control and Prevention, n.d.). So it’s important to analyze the obesity data in the United States.

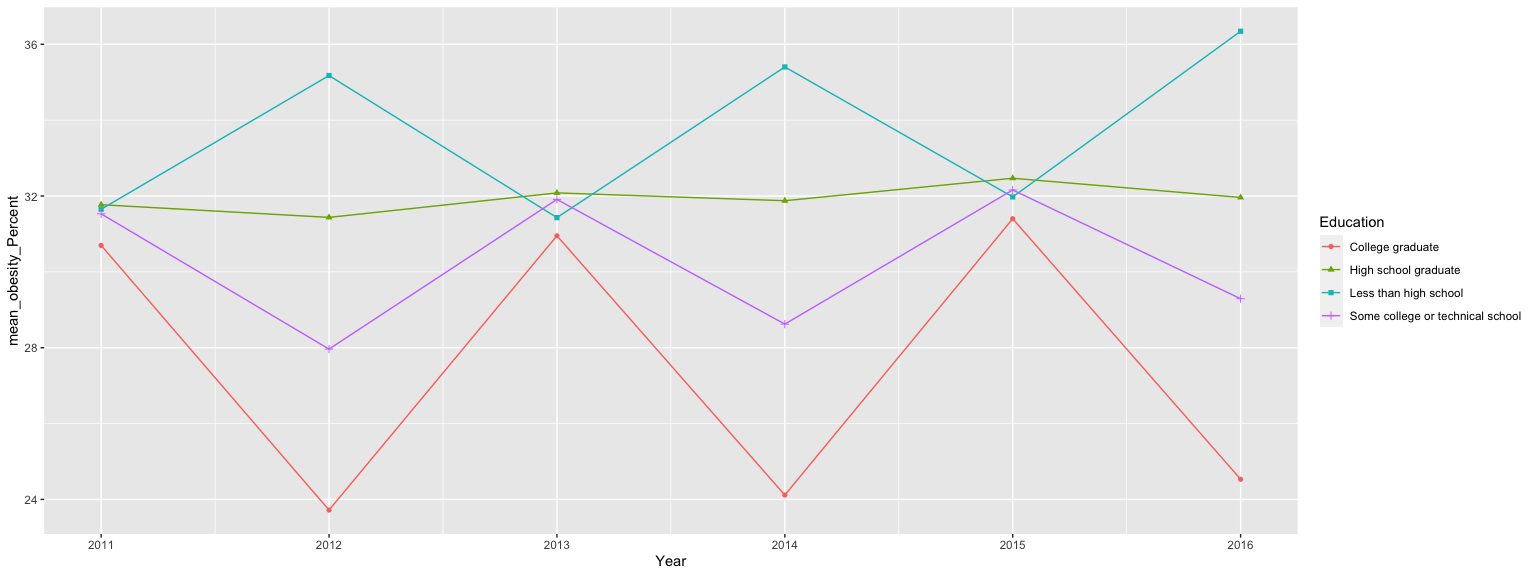
## Data Description

According to the CDC, the prevalence of obesity varies by income or education level, although there may be different patterns in high-income and low-income countries. An analysis of data from the national Health and Nutrition Examination Survey (NHANES) from 2011 to 2014 revealed differences in the relationship between obesity and education, between obesity and income among US adults, and in patterns of obesity epidemics among different races and Hispanics. The prevalence of obesity in women decreased with increasing income (from 45.2% to 29.7%), but there was no difference between the lowest income group (31.5%) and the highest income group (32.6%) among men. The relationship between obesity and income or education level is complex and varies by gender, race or non-Hispanic origin. Let’s take a closer look at how these factors differ depending on how obese the researchers are.

### Income

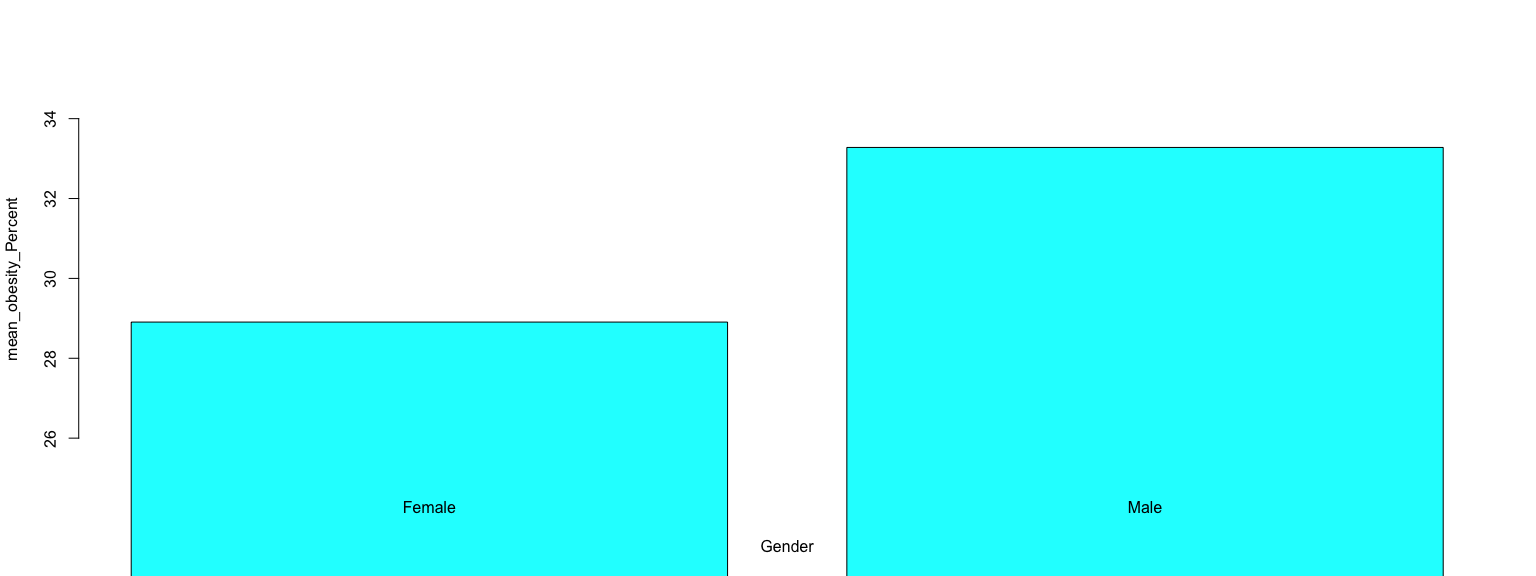
 As can be seen from the figure above, first of all, obesity at all income levels increases as a whole with each year. Secondly, it can be seen that the group with the lowest obesity rate is the group with the highest income, and the group with the highest obesity rate is the group with the lowest income. Obesity rates are significantly inversely proportional to income.

### Education



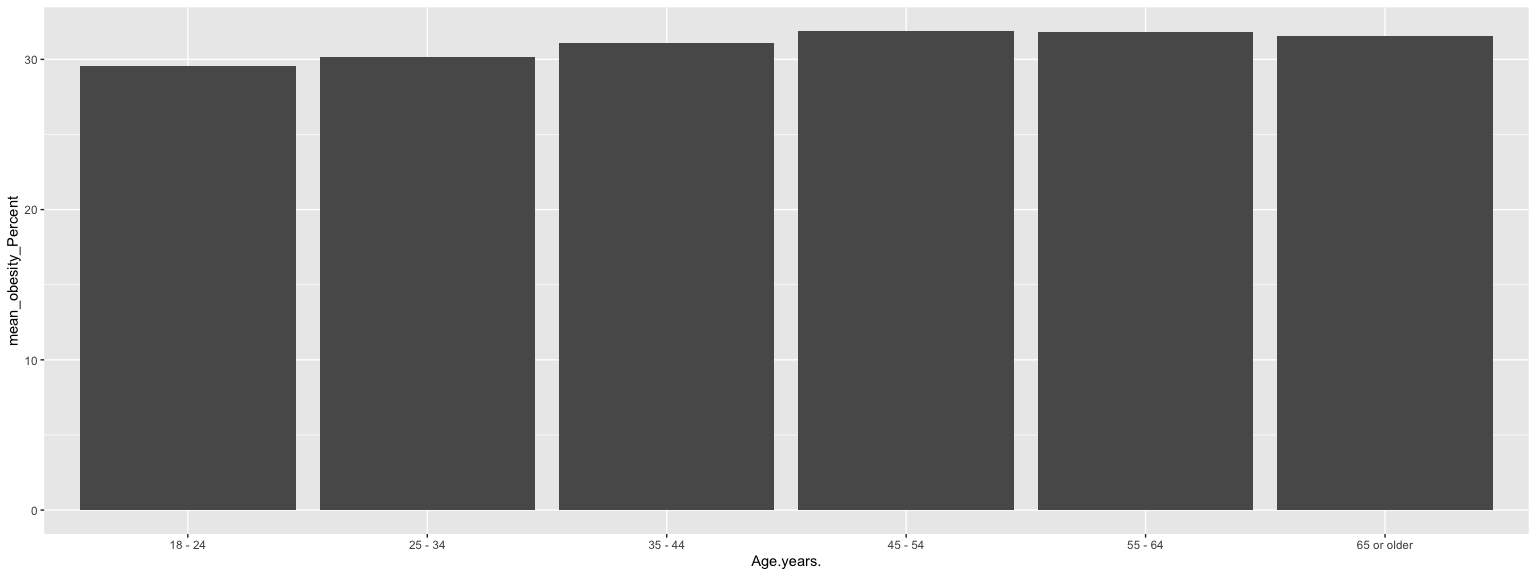
As can be seen from the figure above, first of all, obesity at all levels of education increases with each year as a whole. Secondly, it can be seen that the group with the lowest obesity rate is the group with the highest education level, while the group with the highest obesity rate is the group with the lowest education level. Obesity rates are significantly inversely proportional to educational levels.

### Gender



As you can see, men have significantly higher rates of obesity than women.

### Age



Next, when we look at the age variable, we see that the older you are, the higher the obesity rate.

## Methods

Linear regression is one of the most well-known modeling techniques. It is often one of the first techniques people use when learning predictive models. In this technique, dependent variables are continuous, independent variables can be continuous or discrete, and regression lines are linear in nature. Linear regression establishes a relationship between the dependent variable (Y) and one or more independent variables (X) using the best fitting line (i.e., the regression line). Multiple linear regression can be expressed as Y=a+ B1 *X + B2* X2+ E, where a is the intercept, B is the slope of the line, and E is the error term. Multiple linear regression can predict the value of the target variable based on the given predictive variable. When processing multiple independent variables, stepwise regression is used for variable filtering. In stepwise regression, the selection of independent variables is done in an automatic procedure. It identifies important variables by looking at the values of statistics, such as R-Square, T-Stats, and AIC metrics. Stepwise regression fits the model by simultaneously adding/removing covariables based on specified criteria. The forward selection method and backward elimination method are commonly used stepwise regression methods. The forward selection method starts with the most prominent prediction in the model and adds variables for each step. The backward culling method starts at the same time as all the predictions of the model, and then eliminates the variables of minimum significance at each step. The goal of this modeling technique is to maximize predictive power with a minimum number of predictive variables. This is also one of the ways to deal with high-dimensional data sets.

## Results

##   
## Call:  
## lm(formula = Data\_Value ~ Income, data = new\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -27.967 -6.633 -0.233 5.829 40.929   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.9034 0.2292 139.216 < 2e-16 \*\*\*  
## Income$25,000 - $34,999 -0.4214 0.3241 -1.300 0.19354   
## Income$35,000 - $49,999 -0.5659 0.3241 -1.746 0.08084 .   
## Income$50,000 - $74,999 -0.9704 0.3241 -2.994 0.00276 \*\*   
## Income$75,000 or greater -1.6324 0.3241 -5.037 4.8e-07 \*\*\*  
## IncomeLess than $15,000 0.5637 0.3241 1.739 0.08200 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.01 on 11436 degrees of freedom  
## Multiple R-squared: 0.004798, Adjusted R-squared: 0.004363   
## F-statistic: 11.03 on 5 and 11436 DF, p-value: 1.3e-10

As can be seen, the overall model is significantly established, and income levels do have an impact on obesity rates. In addition, from the perspective of regression coefficients of specific categories, most of the regression coefficients are negative, while p value of the group with the highest income is the smallest, and this category is the most significant. These regression results can further verify our views.

##   
## Call:  
## lm(formula = Data\_Value ~ Education, data = new\_data2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -30.192 -6.952 0.225 5.755 42.325   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 29.2754 0.2409 121.550 < 2e-16  
## EducationHigh school graduate 2.7481 0.3406 8.068 8.24e-16  
## EducationLess than high school 3.4167 0.3406 10.031 < 2e-16  
## EducationSome college or technical school 1.7762 0.3406 5.215 1.89e-07  
##   
## (Intercept) \*\*\*  
## EducationHigh school graduate \*\*\*  
## EducationLess than high school \*\*\*  
## EducationSome college or technical school \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.52 on 7624 degrees of freedom  
## Multiple R-squared: 0.01474, Adjusted R-squared: 0.01435   
## F-statistic: 38.02 on 3 and 7624 DF, p-value: < 2.2e-16

As you can see, the overall model is significant and education does have an impact on obesity rates. In addition, from the perspective of regression coefficients of specific categories, the corresponding regression coefficients of all categories are significant, which can further verify our views.

## Conclusion

As can be seen from the data results, the rate of overweight and obesity of males is higher than that of females. The obesity rate is inversely proportional to education level and income level. The rate of obesity is proportional to age.  
The harm of overweight and obesity is great. Therefore, the treatment of overweight and obesity is of great significance. Ultimately the purpose of treatment to reduce weight, reduce weight can effectively slow down or avoid overweight and obesity on physical and mental damage, the study showed that weight reduction is beneficial to relieve insulin resistance, decrease blood sugar, is good for high blood pressure as well as the improvement of all kinds of metabolic disorders, about 10% of the existing weight, can reduce the risk of a variety of metabolic disorders. Direct cause of overweight and obesity for long-term energy intake than energy expenditure, lead to too much energy in the form of fat storage, the core of diet therapy to restrict energy intake, deployment of diet proportion, advocate science and healthy diet, balanced and nutritious food, and to ensure that the body’s normal energy metabolism and nutrient requirements. In addition, exercise therapy is one of the most effective and basic methods to treat overweight and obesity, which needs to be combined with diet therapy to achieve better weight loss results. The basic principle of exercise therapy is that exercise can make the triglyceride in adipose tissue decompose, release fatty acid, provide energy for the activity of muscle tissue, human body energy to reach the state of balance or negative balance, so as to achieve the goal of weight loss, control overweight, obesity progress.

## Reference

Centers for Disease Control and Prevention. (n.d.). Adult Obesity Causes & Consequences. Centers for Disease Control and Prevention. Retrieved December 5, 2020, from: https://www.cdc.gov/obesity/adult/causes.html

Centers for Disease Control and Prevention. (n.d.). Adult Obesity Facts. Centers for Disease Control and Prevention. Retrieved December 5, 2020, from: https://www.cdc.gov/obesity/data/adult.html

Centers for Disease Control and Prevention. (n.d.). Defining Adult Overweight and Obesity. Centers for Disease Control and Prevention. Retrieved December 5, 2020, from: https://www.cdc.gov/obesity/adult/defining.html

Centers for Disease Control and Prevention. (n.d.). Nutrition, Physical Activity, and Obesity: Data, Trends and Maps. Centers for Disease Control and Prevention. Retrieved December 5, 2020, from: https://nccd.cdc.gov/dnpao\_dtm/rdPage.aspx?rdReport=DNPAO\_DTM.ExploreByLocation&rdRequestForwarding=Form

Ober P B . Introduction to linear regression analysis. journal of the royal statistical society, 2010, 40(12):2775-2776.

Montgomery D C , Peck E A , Vining G G . Introduction to linear regression analysis. Wiley, 1982.

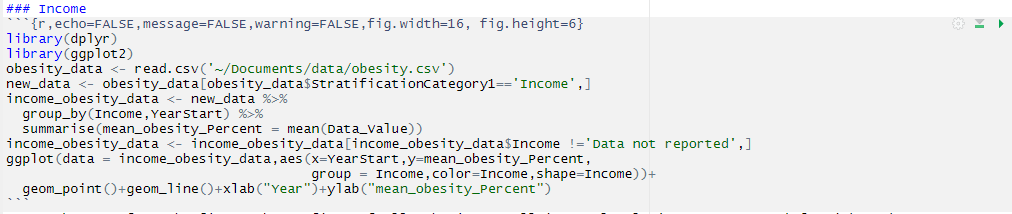
Lockwood, E. H. . (2016). An introduction to linear regression and correlation (2nd edition), by edwards allen l. pp 206. £10·95. 1984. isbn 0-7167-1593-7 1594-5 (pbk) (freeman). Mathematical Gazette, 69(2), 1-17.

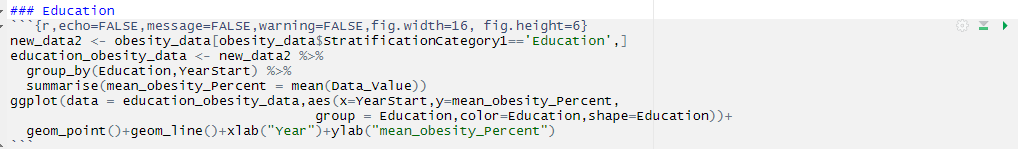
Bramsved, R., Regber, S., Novak, D., Mehlig, K., Lissner, L., & Marild, S. (2018). Parental education and family income affect birthweight, early longitudinal growth, and body mass index development differently. Acta Paediatrica (Oslo, Norway: 1992), 107(11), 1946-1952.

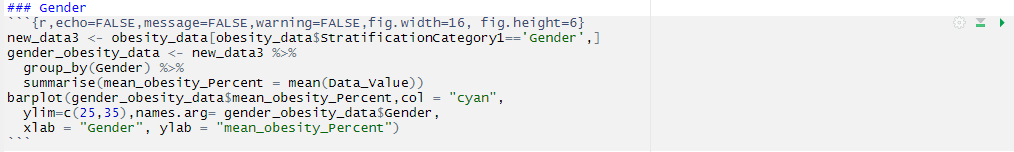
# Appendix

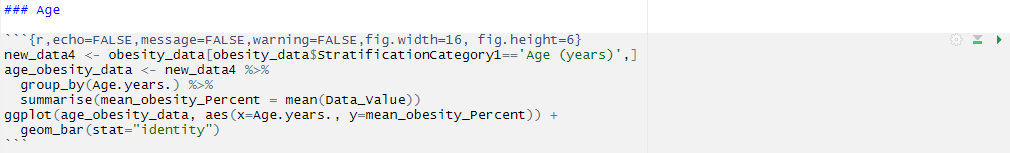
## Appendix A.1 R Code















## Appendix A.2 Data

Since the data id too large, the data can be viewed and download at https://www.kaggle.com/spittman1248/cdc-data-nutrition-physical-activity-obesity