|  |
| --- |
| **Dalhousie University** |
| **STAT 3340** |
| **Regression of Obesity Among Adults** |
|  |
| Group 10  Yutong Liu B00728064  Yifan Chen B00745748  Date: December 11, 2020 |
|  |

## Abstract

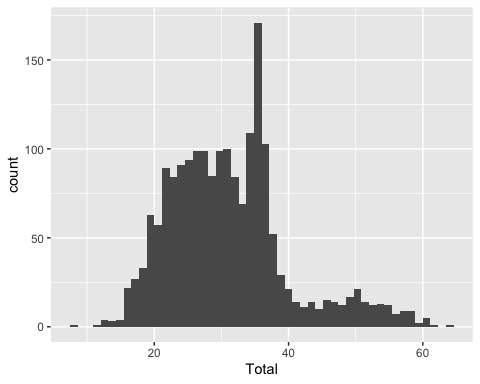
This project is mainly through to the United States in 2011-2016 adults are overweight and obesity of sampling data is analyzed, American adult obesity epidemic characteristics, and related influence factors of overweight and obesity to establish regression model, find out the influence factors of obesity, 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 (Mokdad et al., 2003). It seriously affects the normal physiological function of the body, and has reached the level of damage to health. In 2000, the World Health Organization ranked obesity as one of the top three health killers, along with smoking and AIDS(Yang & Colditz, 2015). 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% (Ogden et al., 2015). 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(Flegal et al., 2012). 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(Ogden et al., 2017). On the basis of the above background, this paper selects four factors of income, age, gender and income to establish a regression model, hoping to see specifically which attributes will influence the obesity rate.

First, look at the distribution of obesity rates as a whole.  
 As can be seen from the histogram above, the overall obesity rate is concentrated in the range of 20-40, with the peak between 35-39.Now let’s look at each of these factors and how they correlate with overall obesity rates.

cor(df1)

## incless15.000 inc15000 inc25000 inc35000 inc50000  
## incless15.000 1.0000000 0.8611806 0.7016771 0.5552332 0.4047288  
## inc15000 0.8611806 1.0000000 0.8642727 0.7693269 0.6455740  
## inc25000 0.7016771 0.8642727 1.0000000 0.8841168 0.8218530  
## inc35000 0.5552332 0.7693269 0.8841168 1.0000000 0.9227319  
## inc50000 0.4047288 0.6455740 0.8218530 0.9227319 1.0000000  
## inc75000 0.1803949 0.4466407 0.6876004 0.8435710 0.9267243  
## age18 0.2227108 0.3326327 0.4785989 0.5792805 0.6348148  
## age25 0.4354457 0.6371096 0.7855681 0.8802761 0.9083685  
## age35 0.5750516 0.7671677 0.8704367 0.9181093 0.9143529  
## age45 0.6624168 0.8264700 0.8942034 0.9170296 0.8854780  
## age55 0.6618508 0.8388705 0.8968013 0.9049336 0.8560989  
## age65 0.5489852 0.7390987 0.8135183 0.8089746 0.7562018  
## college.or.technical 0.5028035 0.7185223 0.8658005 0.9473108 0.9598021  
## College.graduate 0.1455012 0.4108900 0.6568007 0.8193895 0.9086370  
## High.school.graduate 0.8125573 0.9260759 0.9248452 0.8751729 0.7841381  
## Less.than.high.school 0.8922486 0.8480284 0.6542817 0.4798351 0.3051423  
## Male 0.4799725 0.6984414 0.8495589 0.9261068 0.9394747  
## Female 0.6688016 0.8243336 0.8839044 0.8998584 0.8609409  
## Total 0.5962045 0.7950592 0.9079711 0.9577181 0.9457133  
## inc75000 age18 age25 age35 age45  
## incless15.000 0.18039494 0.22271082 0.4354457 0.5750516 0.6624168  
## inc15000 0.44664066 0.33263267 0.6371096 0.7671677 0.8264700  
## inc25000 0.68760043 0.47859895 0.7855681 0.8704367 0.8942034  
## inc35000 0.84357104 0.57928054 0.8802761 0.9181093 0.9170296  
## inc50000 0.92672433 0.63481481 0.9083685 0.9143529 0.8854780  
## inc75000 1.00000000 0.72636604 0.8868232 0.8231952 0.7600776  
## age18 0.72636604 1.00000000 0.7751014 0.5597730 0.4585103  
## age25 0.88682316 0.77510141 1.0000000 0.8831263 0.8160445  
## age35 0.82319524 0.55977304 0.8831263 1.0000000 0.9318293  
## age45 0.76007760 0.45851032 0.8160445 0.9318293 1.0000000  
## age55 0.72600925 0.36796404 0.7474929 0.8895783 0.9498246  
## age65 0.66886339 0.32290584 0.6090152 0.7127152 0.7961198  
## college.or.technical 0.91084975 0.68214248 0.9263974 0.9371676 0.9150098  
## College.graduate 0.98555998 0.73909531 0.8735647 0.7876640 0.7246407  
## High.school.graduate 0.61386206 0.46801113 0.7560187 0.8578399 0.9067027  
## Less.than.high.school 0.06089981 0.04922866 0.3091420 0.4854026 0.5872794  
## Male 0.89547329 0.69663337 0.9267820 0.9191657 0.8873949  
## Female 0.76831770 0.56132357 0.8207410 0.8845061 0.9166950  
## Total 0.87486090 0.65849252 0.9172531 0.9463622 0.9454633  
## age55 age65 college.or.technical College.graduate  
## incless15.000 0.6618508 0.5489852 0.5028035 0.14550116  
## inc15000 0.8388705 0.7390987 0.7185223 0.41088996  
## inc25000 0.8968013 0.8135183 0.8658005 0.65680069  
## inc35000 0.9049336 0.8089746 0.9473108 0.81938952  
## inc50000 0.8560989 0.7562018 0.9598021 0.90863699  
## inc75000 0.7260092 0.6688634 0.9108497 0.98555998  
## age18 0.3679640 0.3229058 0.6821425 0.73909531  
## age25 0.7474929 0.6090152 0.9263974 0.87356472  
## age35 0.8895783 0.7127152 0.9371676 0.78766397  
## age45 0.9498246 0.7961198 0.9150098 0.72464074  
## age55 1.0000000 0.8791819 0.8844079 0.69470730  
## age65 0.8791819 1.0000000 0.7680011 0.66682899  
## college.or.technical 0.8844079 0.7680011 1.0000000 0.88600802  
## College.graduate 0.6947073 0.6668290 0.8860080 1.00000000  
## High.school.graduate 0.9027580 0.8063325 0.8335399 0.57306049  
## Less.than.high.school 0.6098440 0.5465468 0.3782688 0.02120848  
## Male 0.8470520 0.7479405 0.9443967 0.86810955  
## Female 0.9212107 0.8496094 0.9149277 0.75974032  
## Total 0.9261440 0.8366424 0.9747383 0.85540942  
## High.school.graduate Less.than.high.school Male  
## incless15.000 0.8125573 0.89224864 0.4799725  
## inc15000 0.9260759 0.84802843 0.6984414  
## inc25000 0.9248452 0.65428173 0.8495589  
## inc35000 0.8751729 0.47983508 0.9261068  
## inc50000 0.7841381 0.30514233 0.9394747  
## inc75000 0.6138621 0.06089981 0.8954733  
## age18 0.4680111 0.04922866 0.6966334  
## age25 0.7560187 0.30914196 0.9267820  
## age35 0.8578399 0.48540262 0.9191657  
## age45 0.9067027 0.58727938 0.8873949  
## age55 0.9027580 0.60984404 0.8470520  
## age65 0.8063325 0.54654677 0.7479405  
## college.or.technical 0.8335399 0.37826875 0.9443967  
## College.graduate 0.5730605 0.02120848 0.8681096  
## High.school.graduate 1.0000000 0.74886060 0.8412696  
## Less.than.high.school 0.7488606 1.00000000 0.4029046  
## Male 0.8412696 0.40290460 1.0000000  
## Female 0.8789405 0.55631586 0.8208428  
## Total 0.9007961 0.49971540 0.9591332  
## Female Total  
## incless15.000 0.6688016 0.5962045  
## inc15000 0.8243336 0.7950592  
## inc25000 0.8839044 0.9079711  
## inc35000 0.8998584 0.9577181  
## inc50000 0.8609409 0.9457133  
## inc75000 0.7683177 0.8748609  
## age18 0.5613236 0.6584925  
## age25 0.8207410 0.9172531  
## age35 0.8845061 0.9463622  
## age45 0.9166950 0.9454633  
## age55 0.9212107 0.9261440  
## age65 0.8496094 0.8366424  
## college.or.technical 0.9149277 0.9747383  
## College.graduate 0.7597403 0.8554094  
## High.school.graduate 0.8789405 0.9007961  
## Less.than.high.school 0.5563159 0.4997154  
## Male 0.8208428 0.9591332  
## Female 1.0000000 0.9487543  
## Total 0.9487543 1.0000000

It can be seen from the correlation coefficient that, on the whole, the correlation coefficient between all variables is large.

## 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 (Ober ,2010).

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(Montgomery et al., 1982). 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 datasets (Lockwood,2016).

## Results

##   
## Call:  
## lm(formula = Total ~ ., data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.35032 -0.05245 -0.00524 0.04534 0.48868   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0282136 0.0088482 -3.189 0.001453 \*\*   
## incless15.000 -0.0039584 0.0006965 -5.683 1.53e-08 \*\*\*  
## inc15000 0.0002327 0.0009177 0.254 0.799819   
## inc25000 -0.0011842 0.0007468 -1.586 0.112951   
## inc35000 -0.0042950 0.0008794 -4.884 1.13e-06 \*\*\*  
## inc50000 -0.0022028 0.0009286 -2.372 0.017778 \*   
## inc75000 -0.0026585 0.0012103 -2.197 0.028173 \*   
## age18 0.0036900 0.0009921 3.719 0.000206 \*\*\*  
## age25 0.0128458 0.0014431 8.901 < 2e-16 \*\*\*  
## age35 0.0127892 0.0013805 9.264 < 2e-16 \*\*\*  
## age45 0.0132978 0.0015134 8.787 < 2e-16 \*\*\*  
## age55 0.0108412 0.0014804 7.323 3.57e-13 \*\*\*  
## age65 0.0163491 0.0014394 11.358 < 2e-16 \*\*\*  
## college.or.technical 0.0059182 0.0015265 3.877 0.000109 \*\*\*  
## College.graduate 0.0073107 0.0014237 5.135 3.11e-07 \*\*\*  
## High.school.graduate 0.0062606 0.0014477 4.324 1.61e-05 \*\*\*  
## Less.than.high.school 0.0041113 0.0007407 5.551 3.25e-08 \*\*\*  
## Male 0.4593186 0.0035854 128.107 < 2e-16 \*\*\*  
## Female 0.4618379 0.0036034 128.166 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08893 on 1887 degrees of freedom  
## Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999   
## F-statistic: 1.084e+06 on 18 and 1887 DF, p-value: < 2.2e-16

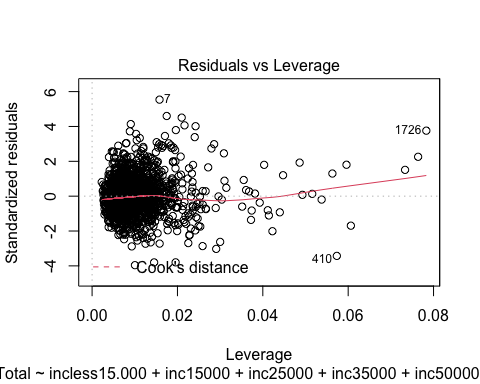
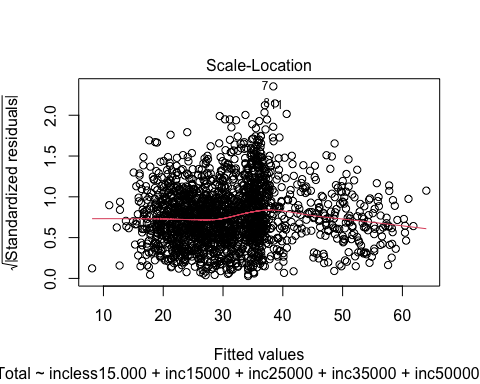
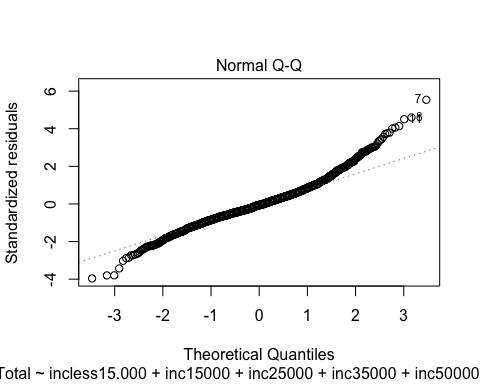
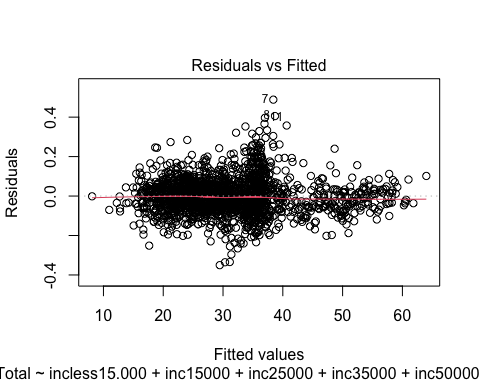
As can be seen from the full model, most of the variables are significant, but several categories of income are not. Next, linear model selection is carried out through forward regression.

## Start: AIC=-9205.79  
## Total ~ incless15.000 + inc15000 + inc25000 + inc35000 + inc50000 +   
## inc75000 + age18 + age25 + age35 + age45 + age55 + age65 +   
## college.or.technical + College.graduate + High.school.graduate +   
## Less.than.high.school + Male + Female

##   
## Call:  
## lm(formula = Total ~ incless15.000 + inc15000 + inc25000 + inc35000 +   
## inc50000 + inc75000 + age18 + age25 + age35 + age45 + age55 +   
## age65 + college.or.technical + College.graduate + High.school.graduate +   
## Less.than.high.school + Male + Female, data = df1)  
##   
## Coefficients:  
## (Intercept) incless15.000 inc15000   
## -0.0282136 -0.0039584 0.0002327   
## inc25000 inc35000 inc50000   
## -0.0011842 -0.0042950 -0.0022028   
## inc75000 age18 age25   
## -0.0026585 0.0036900 0.0128458   
## age35 age45 age55   
## 0.0127892 0.0132978 0.0108412   
## age65 college.or.technical College.graduate   
## 0.0163491 0.0059182 0.0073107   
## High.school.graduate Less.than.high.school Male   
## 0.0062606 0.0041113 0.4593186   
## Female   
## 0.4618379

final\_model <- lm(formula = Total ~ incless15.000 + inc15000 + inc25000 + inc35000 +   
 inc50000 + inc75000 + age18 + age25 + age35 + age45 + age55 +   
 age65 + college.or.technical + College.graduate + High.school.graduate +   
 Less.than.high.school + Male + Female, data = df1)  
summary(final\_model)

##   
## Call:  
## lm(formula = Total ~ incless15.000 + inc15000 + inc25000 + inc35000 +   
## inc50000 + inc75000 + age18 + age25 + age35 + age45 + age55 +   
## age65 + college.or.technical + College.graduate + High.school.graduate +   
## Less.than.high.school + Male + Female, data = df1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.35032 -0.05245 -0.00524 0.04534 0.48868   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0282136 0.0088482 -3.189 0.001453 \*\*   
## incless15.000 -0.0039584 0.0006965 -5.683 1.53e-08 \*\*\*  
## inc15000 0.0002327 0.0009177 0.254 0.799819   
## inc25000 -0.0011842 0.0007468 -1.586 0.112951   
## inc35000 -0.0042950 0.0008794 -4.884 1.13e-06 \*\*\*  
## inc50000 -0.0022028 0.0009286 -2.372 0.017778 \*   
## inc75000 -0.0026585 0.0012103 -2.197 0.028173 \*   
## age18 0.0036900 0.0009921 3.719 0.000206 \*\*\*  
## age25 0.0128458 0.0014431 8.901 < 2e-16 \*\*\*  
## age35 0.0127892 0.0013805 9.264 < 2e-16 \*\*\*  
## age45 0.0132978 0.0015134 8.787 < 2e-16 \*\*\*  
## age55 0.0108412 0.0014804 7.323 3.57e-13 \*\*\*  
## age65 0.0163491 0.0014394 11.358 < 2e-16 \*\*\*  
## college.or.technical 0.0059182 0.0015265 3.877 0.000109 \*\*\*  
## College.graduate 0.0073107 0.0014237 5.135 3.11e-07 \*\*\*  
## High.school.graduate 0.0062606 0.0014477 4.324 1.61e-05 \*\*\*  
## Less.than.high.school 0.0041113 0.0007407 5.551 3.25e-08 \*\*\*  
## Male 0.4593186 0.0035854 128.107 < 2e-16 \*\*\*  
## Female 0.4618379 0.0036034 128.166 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.08893 on 1887 degrees of freedom  
## Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999   
## F-statistic: 1.084e+06 on 18 and 1887 DF, p-value: < 2.2e-16

plot(final\_model)

As can be seen from the diagnostic diagram, a regression model selected by step forward is reasonable.

To sum up, the age, income, education level and gender of the people we choose also have an impact on obesity rates. In addition, in terms of the regression coefficients of specific categories, the corresponding regression coefficients of each category are basically significant.

## Conclusion

As can be seen from the overall regression analysis, the income is inversely proportional to the obesity rate, and the higher the income, the lower the obesity rate. It can also be seen in the regression coefficient comparison of the various categories of age that the higher the age, the greater the regression coefficient and the age is proportional to the obesity rate. The results of regression analysis have some implications for reducing obesity rate. We should pay more attention to the obesity rate among low-income and elderly people. It is interesting to guide the health awareness of these people and organize them to participate in sports. Exercise therapy is one of the most effective and basic treatments for overweight and obesity. It can also lead to healthy eating habits. Diet needs to be combined with diet therapy to achieve better weight loss results.

## Reference

Mokdad, A. H. , Ford, E. S. , Bowman, B. A. , Dietz, W. H. , & Marks, J. S. . (2003). Prevalence of obesity, diabetes, and obesity-related health risk factors,2001. *Jama the Journal of the American Medical Association, 289*(1), 76-9.

Yang, L. , & Colditz, G. A. . (2015). Prevalence of overweight and obesity in the united states, 2007-2012. *Jama Internal Medicine, 175*(8), 1412.

Ogden, C. L. , Carroll, M. D. , Kit, B. K. , & Flegal, K. M. . (2014). Prevalence of childhood and adult obesity in the united states, 2011-2012. *Jama, 311*(8), 806.

Flegal, K. M. , Carroll, M. D. , Kit, B. K. , & Ogden, C. L. . (2012). Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999-2010. *Jama the Journal of the American Medical Association, 307*(5), 483-90.

Ogden CL, Fakhouri TH, Carroll MD, et al. (2017). Prevalence of Obesity Among Adults, by Household Income and Education — United States, 2011–2014. *MMWR Morb Mortal Wkly Rep 2017*;66:1369–1373. DOI: http://dx.doi.org/10.15585/mmwr.mm6650a1

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

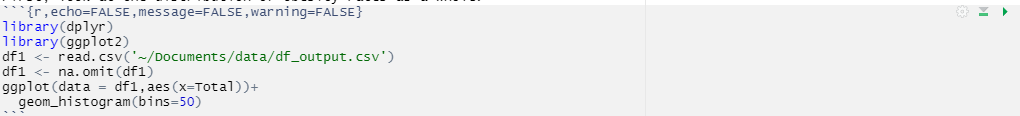
Douglas C.. Montgomery, Peck, E. A., & Vining, G. G. (2001). *Introduction to linear regression analysis*. Wiley.

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.

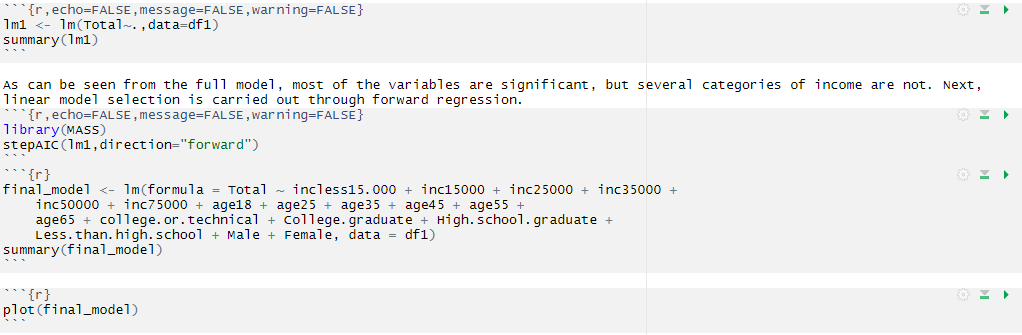
# Appendix

## Appendix 1. R Code









## Appendix 2. Data

