Lifestyle and Adult Obesity

490241543

Introduction

People who are overweight likely to carry a higher risk to be exposed to unhealthy conditions and many critical diseases, such as hypertension, dyslipidemia, diabetes, stroke, osteoarthritis, breathing problem, mental disorders, and sleep apnea. The obesity may cause all-causes of mortality and result in low quality of life in terms of both mental and physical. Even though the quality of life is a quite young field of research, dozens of researchers demonstrated that obesity is an underlying factor in terms of reduced quality of life, compared with normal weight.

This research mainly aims to find what factors really affect to adult's obesity rate at the county level in the United States. Obesity is a complex subject with multiple causes, but the people's lifestyle has received the most attention than epidemiology trends (e.g. gender, race, ethnicity, etc.) or genetic factors in this study. Here, lifestyle is justified as the accessibility to healthy food and the opportunity of exercise, but some additional data such as income that might impact the choice of the lifestyle is also used. Moreover, I hypothesise that people who have more frequent physically inactivate days and consume more low quality of food are more likely to be overweight.

Diverse types of visualisation and analysis will be stated for comparison with my theory and to find out affective factors of the obesity prevalence through this paper.

Literature and theory

In order to support the hypotheses that I came up with selecting the attributes to be used for this study, some literature review had to be done as follows.

The accessibility to healthy food

Many studies have successfully examined the link between healthy food and obesity or Body Mass Index(BMI). First of all, a recent study on obesity and the food environment shows that better access to a supermarket reduces the chance of obesity and better access to a convenient store raises the chance of obesity[1]. Moreover, one literature empirically proved the correlation between the distance to the supermarket from home and the healthy produce, suggesting that participants with no supermarkets within 1 mile of their home (the lowest category in the survey) were 25 percent less likely to have a healthy diet than participants who had the most stores near their home[2]. It is also noticeable that economically deprived neighbourhoods (counties) tend to have fewer local supermarket carrying fresh products such as vegetables and fruits, but have more fast food retails and convenient stores. Besides, lower income individuals are apt to consume more sugar-sweetened beverages, energy-dense food, and have relatively lower diet quality in overall[3]. On the other hand, one study discovered increasing obesity prevalence along with decreasing fruit and vegetable consumption with increasing distance to supermarkets is found in metropolitan areas, but not significant in other areas[4]. This might limit showing the strong relationship between the obesity rate and healthy food consumption.

The accessibility to the opportunity of exercise

The fact that regular physical activity is considered as an effective preventer of such diseases is irrefutable[5], and physical inactivity is a contributor along with excess caloric intake to weight gain and obesity[6]. Thus, some study suggests creating a healthy environment, including infrastructure, and other physical elements (e.g., sidewalks, bicycle lanes, parks, playgrounds) as a strategy for increasing physical activity[7]. Such kind of healthy-friendly environment is relatively more well organised and kept aesthetic in the metropolitan areas compared to the rural area. As rural areas lack exercise facilities and infrastructure and rural residents report less physical activity, the obesity prevalence rate among rural Americans tends to be higher than their urban counterparts[8]. In addition, since rural households generally have a higher median income than urban households, the correlation between income and physical activeness was also analysed by the number of literature. A number of studies came across that deprivation and poverty were found to be associated with low levels of physical activity[9], and one paper also demonstrated that the high-income strata were 1.9 times more likely to meet physical activity guidelines during a 7-day period than their lower income counterparts[10].

Data and methodology

'2018 County Health Rankings Data' describes a rich variety of health outcomes and potential factors. As this study narrows down the topic to adult obesity, only several attributes, that are related to the subject directly or indirectly, are selected for analysis. External resources or detailed information of the data are also collected from County Health Rankings[11] and Centers for Diseases Control and Prevention[12] website. More descriptions can be found below.

Independent Variable

Food Insecurity : Percentage of the population who did not have access to a reliable source of food during the past year.

Limited Access to Healthy Foods : Percentage of the population that does not live close to a grocery store (Different definition of living close to a grocery store - less than 10 miles in rural areas, less than one mile in non-rural areas).

The variables mentioned above are expected to show evidence that food desert areas are correlated with a high prevalence of obesity.

Physical Inactivity : Percentage of adults reporting no leisure-time physical activity in the past months. As mentioned earlier, physical inactivity is one of the leading contributors to obesity.

Exercise Access : Percentage of individuals in a county who live reasonably close to a location for physical activity. This attribute is expected to show the relationship between the accessibility to the exercise facilities and adult obesity.

Frequent Physical Distress : Percentage of adults reporting 14 or more days of poor physical health per month emphasising more chronic physical issues.

Poor Physical Health Days : Average number of physically unhealthy days reported in the past 30 days. Last two variables were chosen because physically inactivate people are more likely to suffer more physically unhealthy days.

Median Household Income : Income is one the most leading factors of providing options for healthy choices of lifestyle. Poor families or individuals are most likely to be exposed to limited access to healthy foods and physical activities.

Dependent Variables

Adult Obesity : Percentage of the adult population that reports a body mass index (BMI) greater than or equal to 30kg/m2, which was coordinated by United States Diabetes Surveillance System. This dataset was selected as a health outcome on county-level.

Limitations

1. While 'Food Insecurity' records a past few years, 'Physical Inactivity' records a past few months. This causes inconsistent data collection period.
2. 'Access to Exercise Opportunities' is not inclusive of all opportunities within a community. For example, the gyms and private recreational studios might have not been captured in this measurement.
3. Some variables are biased. For instance, 'Limited access to healthy food' and access to 'Frequent Physical Distress' data show left-skewed distribution.

Methodology

Factor Analysis : In this study, factor analysis was used to reduce the number of variables with measuring similarity. By doing factor analysis, the latent trait was estimated to be used later for fitting the regression model.

Multiple Linear Regression : Linear regression conducts a linear estimation of the relationship between a dependent variable and the independent variable(s), and a linear relationship is usually considered as the presumption in social science. Looking at the graphs illustrating the relationship between the dependent variable (Adult obesity) and independent variables below, you can see that they show a reasonable linear relationship. Furthermore, the latent trait of physical data that was obtained from factor analysis is linearly correlated with adult obesity. Thus, multiple linear regression is used to examine the factors for better health outcome (lower adult obesity rate) in this study.

Loading Data

```
## New names:
## * `95% CI - Low` -> `95% CI - Low...5`
## * `95% CI - High` -> `95% CI - High...6`
## * Quartile -> Quartile...7
## * `95% CI - Low` -> `95% CI - Low...12`
## * `95% CI - High` -> `95% CI - High...13`
## * ... and 76 more problems
```

```
## New names:
## * `# Deaths` -> `# Deaths...4`
## * `95% CI - Low` -> `95% CI - Low...6`
## * `95% CI - High` -> `95% CI - High...7`
## * `# Deaths` -> `# Deaths...11`
## * `95% CI - Low` -> `95% CI - Low...13`
## * ... and 30 more problems
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

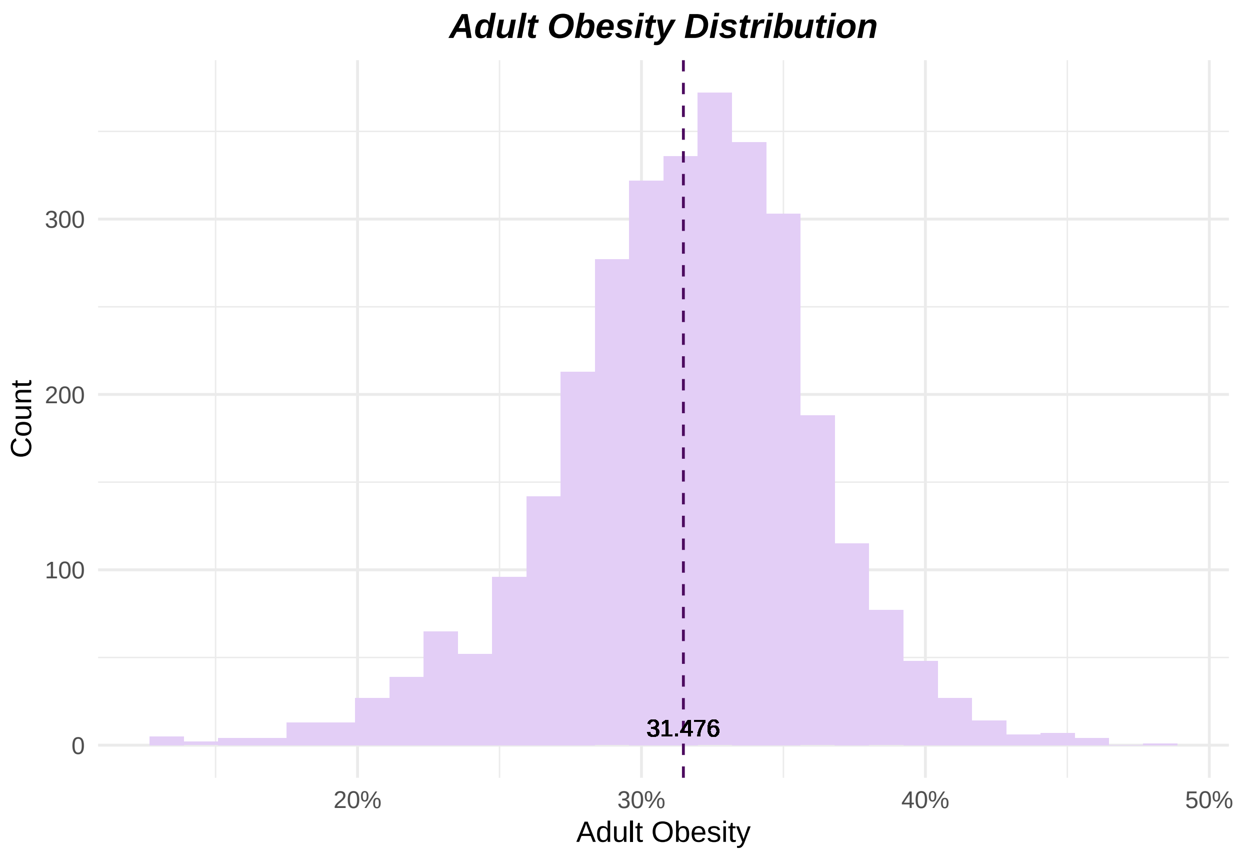
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Loading required package: magrittr
```

Results

Dependent Variable Analysis

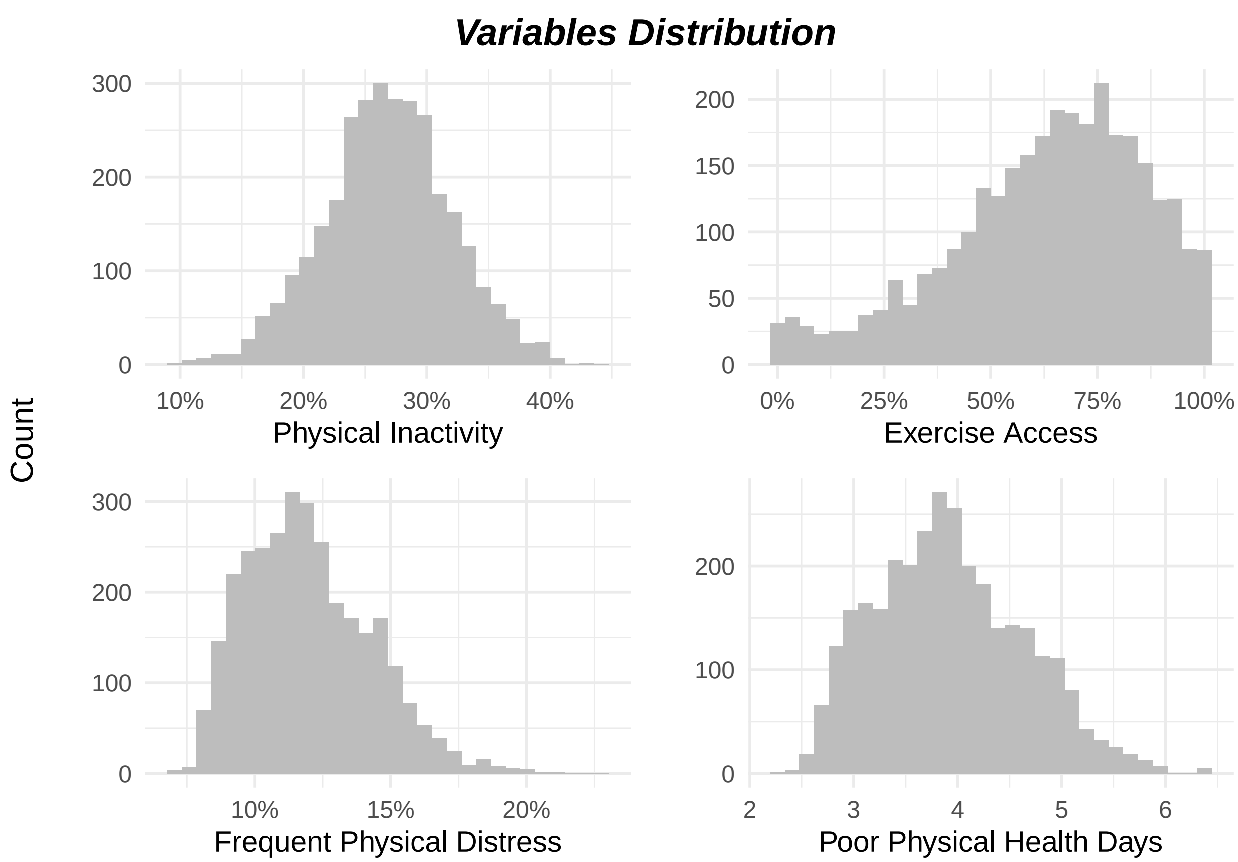
Looking at the histogram below, Adult obesity is normally distributed with a mean value of 31.47648%.

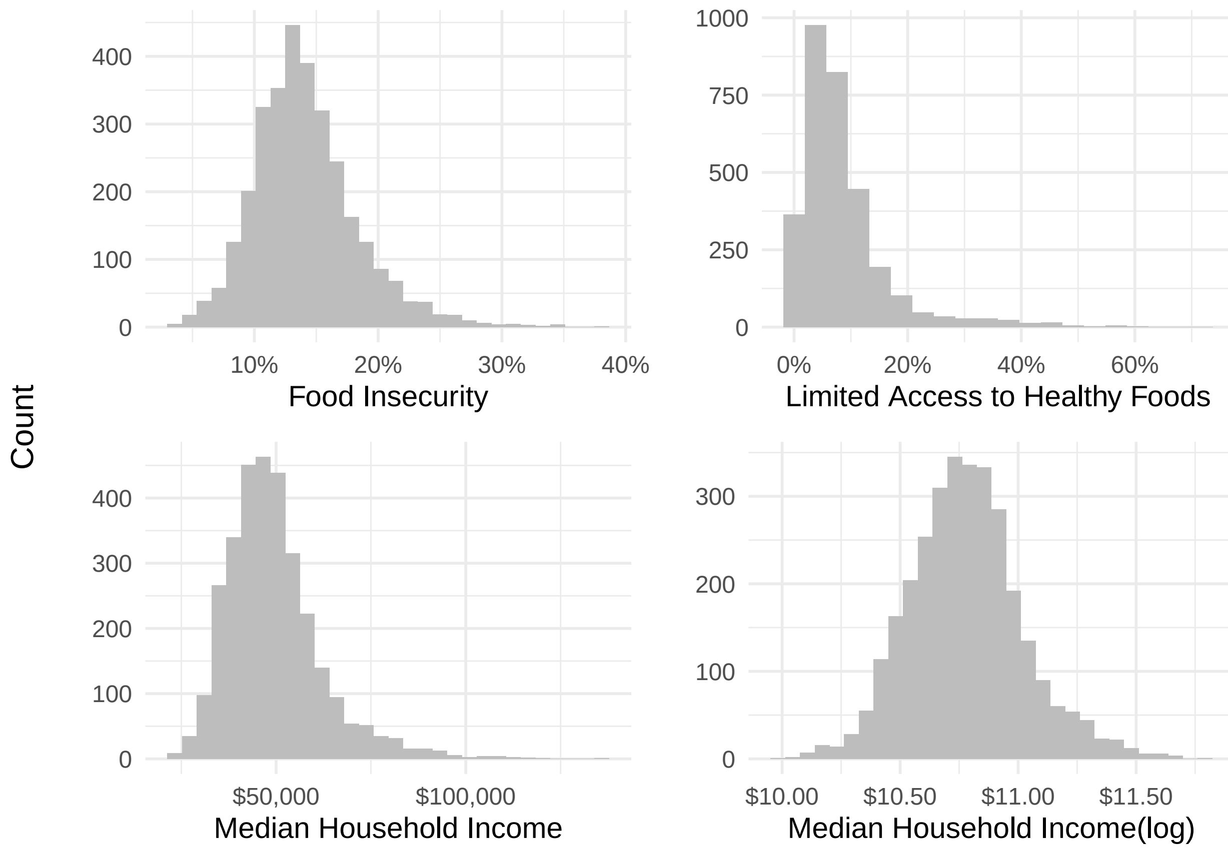


Independent Variable Analysis

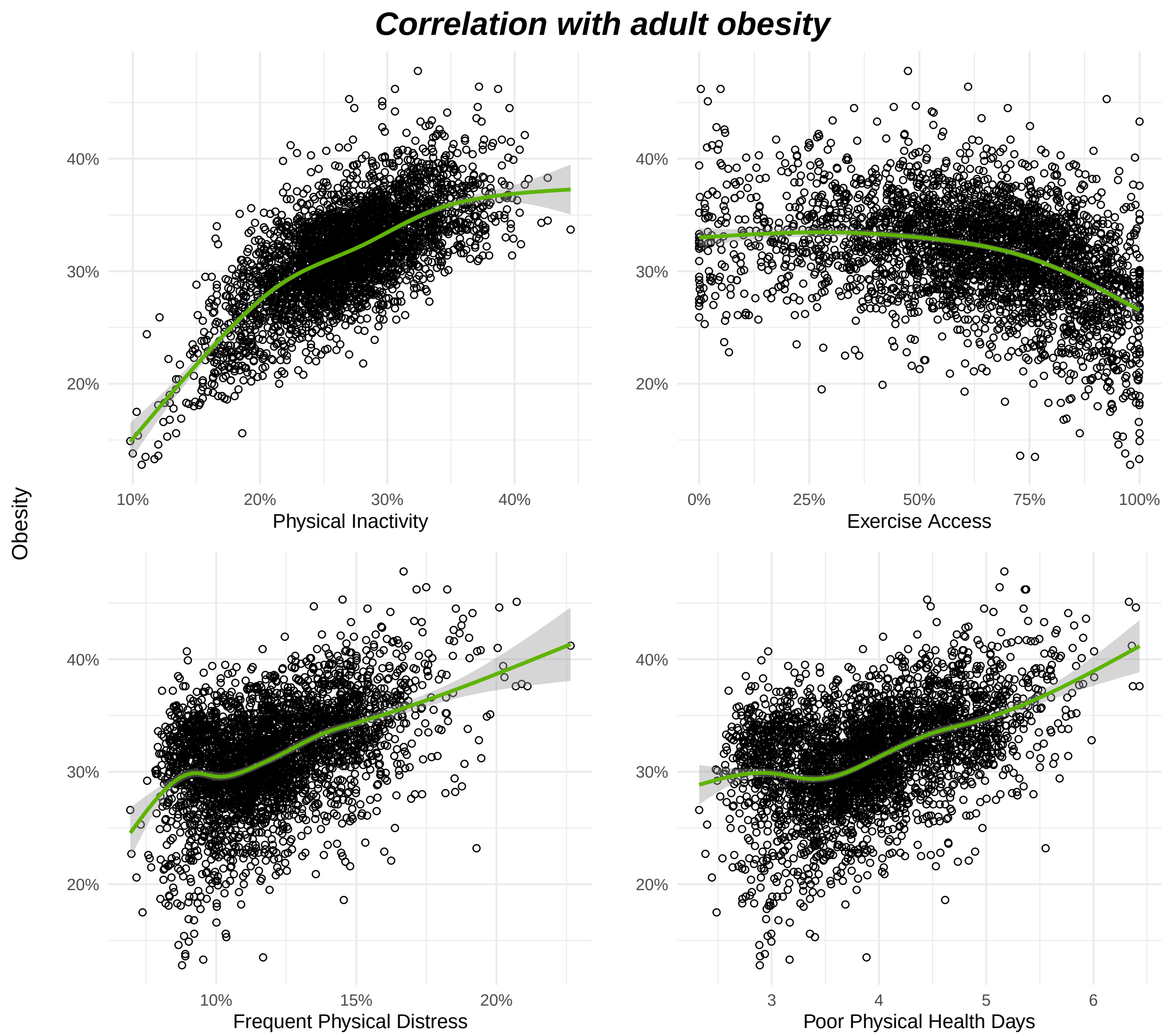
The histograms below show how independent variables are distributed. First of all, physical inactivity is ideally distributed. However, exercise access and frequent physical distress are right-skewed and left-skewed, respectively. Poor physical health data is fairly normally distributed, but slightly biased to the left. Food insecurity variable shows a normal distribution with an insignificant tail. Whereas, limited access to healthy food is extremely left-skewed since the majority of the samples stays under 10%. Lastly, median household income has a very wide range and also biased. Hence, the logarithm transformation was conducted on the income variable. As you can see from the last graph, after income has converted into log, skewness of distribution seems to be corrected and the range of variable has also been narrowed.

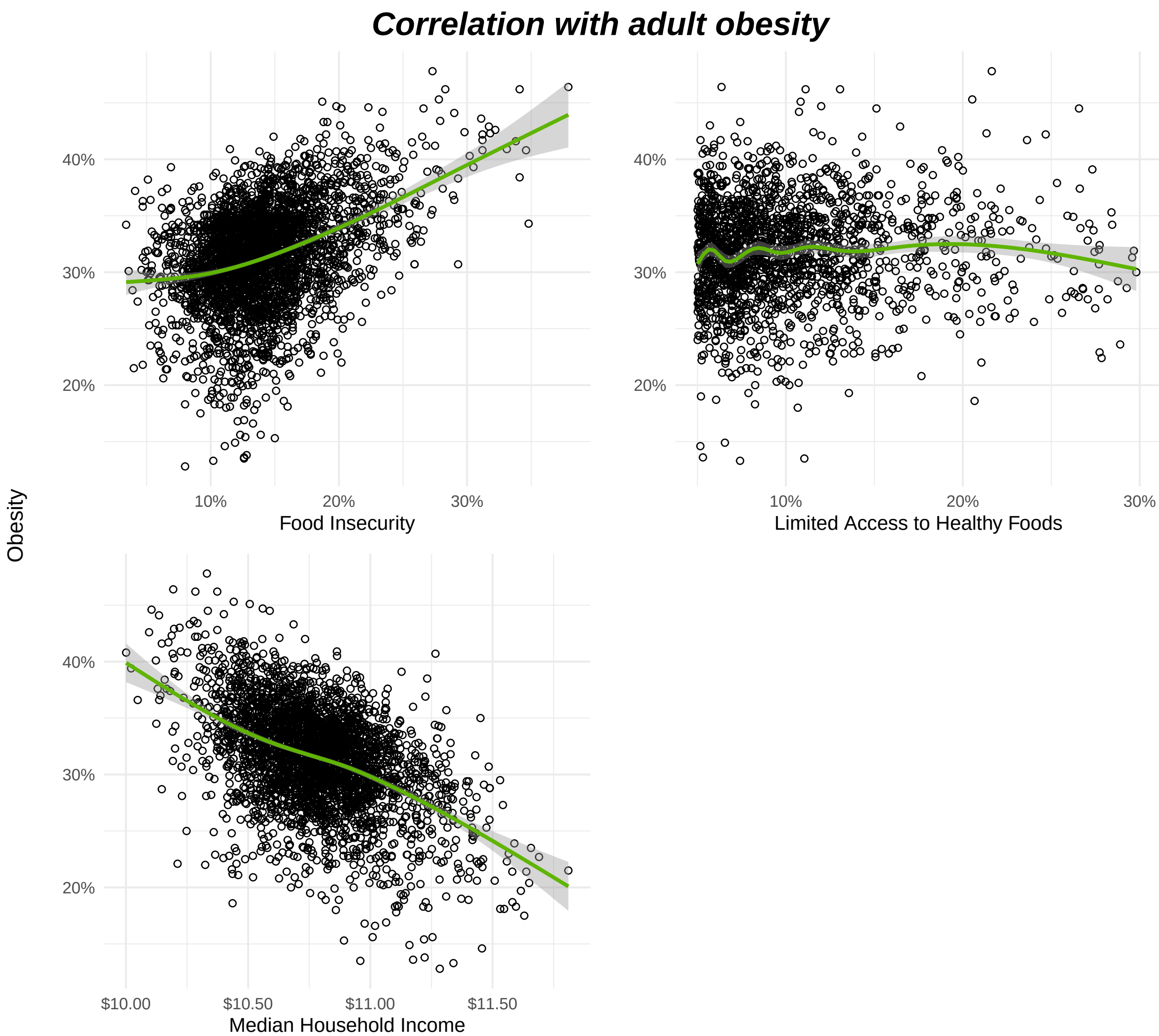
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```






```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```


Correlation with adult obesity
Obesity
40%
30%
20%
10%
20%
30%
40%
Physical Inactivity
0%
25%
50%
75%
100%
Exercise Access
10%
15%
20%
Frequent Physical Distress
3
4
5
6
Poor Physical Health Days

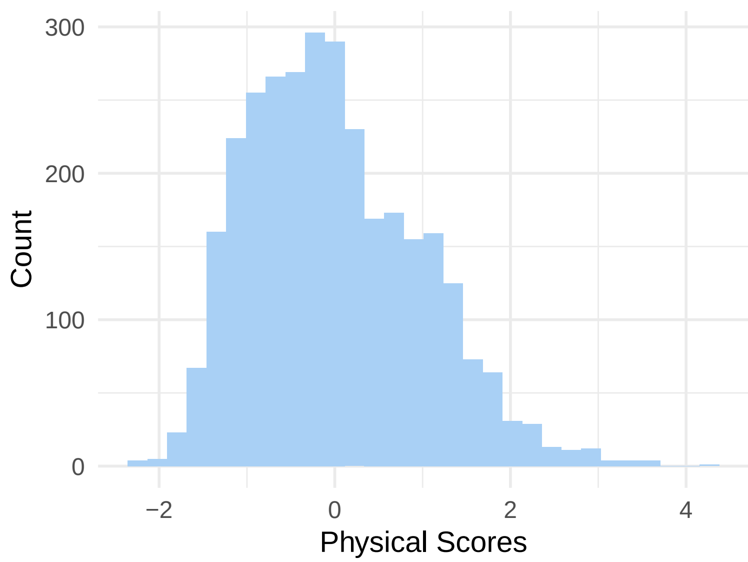

Correlation with adult obesity
Obesity
40%
30%
20%
10%
20%
30%
Food Insecurity
40%
30%
20%
10%
20%
30%
Limited Access to Healthy Foods
40%
30%
20%
$10.00
$10.50
$11.00
$11.50
Median Household Income

Factor Analysis

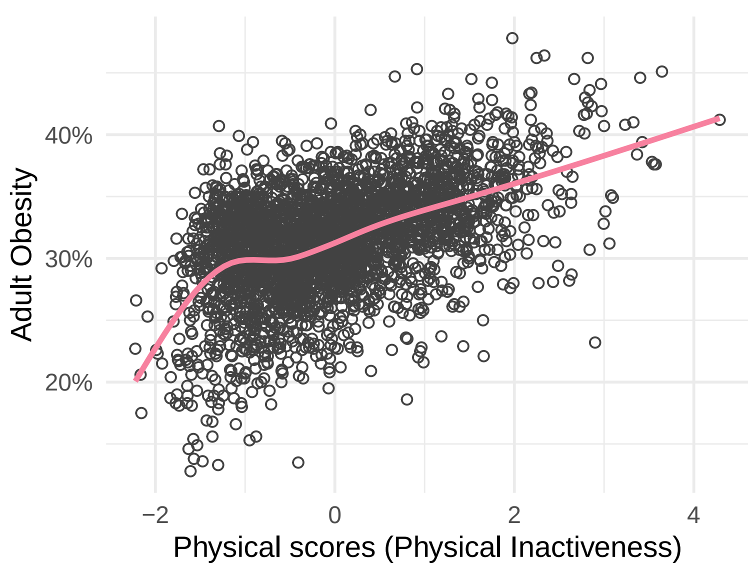
```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:scales':
##
##     alpha, rescale

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

## Factor Analysis using method =  minres
## Call: fa(r = data[, c("physically.inactivate", "exercise.access", "physical.distress",
##     "physically.unhealthy")], nfactors = 1)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                        MR1   h2    u2 com
## physically.inactivate  0.58 0.34 0.663   1
## exercise.access       -0.38 0.14 0.858   1
## physical.distress      0.97 0.95 0.054   1
## physically.unhealthy   0.95 0.91 0.088   1
##
##                 MR1
## SS loadings    2.34
## Proportion Var 0.58
##
## Mean item complexity =  1
## Test of the hypothesis that 1 factor is sufficient.
##
## The degrees of freedom for the null model are  6  and the objective function was  3.9 with Chi Square
## The degrees of freedom for the model are 2  and the objective function was  0.8
##
## The root mean square of the residuals (RMSR) is  0.11
## The df corrected root mean square of the residuals is  0.18
##
## The harmonic number of observations is  3116 with the empirical chi square  420.92  with prob <  4e-9
## The total number of observations was  3116  with Likelihood Chi Square =  2475.2  with prob <  0
##
## Tucker Lewis Index of factoring reliability =  0.389
## RMSEA index =  0.63  and the 90 % confidence intervals are  0.609 0.651
## BIC =  2459.11
## Fit based upon off diagonal values = 0.96
## Measures of factor score adequacy
##                                                   MR1
## Correlation of (regression) scores with factors   0.98
## Multiple R square of scores with factors          0.96
## Minimum correlation of possible factor scores     0.92
```

As shown in the code above, four different variables ("physically.inactivate", "exercise.access", "physical.distress", "physically.unhealthy") imply physical-related information. Looking at the MR1 output that shows how each attribute loads onto the latent trait for physical inactiveness, three items are positively loaded while one item is negatively loaded. What it represents is that higher response of exercise access predicts higher activeness. As I hypothesised, people who are more physically inactive, have more frequent poor physical health days, suffer from more chronic physical issues, and have fewer chances to access to the exercise in their county tend to remarkably affect to the physical inactiveness. The second column, h2 indicates the communality of each item and a higher communality (close to 1) means that an item reflects the common factor better. According to the result, exercise access and physical inactivity show lower h2 values compared to the other two variables that are over 0.9.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



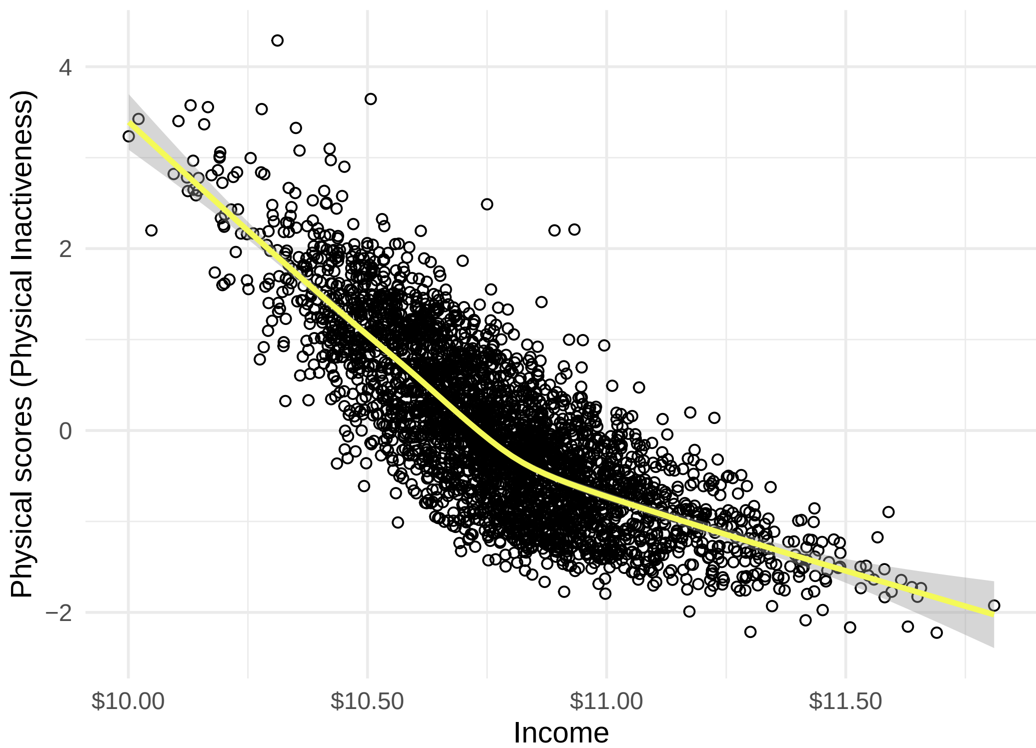
Linear Regression

```
##
## Call:
## lm(formula = obesity ~ physical.scores + food.insecure + food.limited +
##     income.log, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.5311  -2.3245   0.2731   2.7075  12.9422
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     71.343507   5.349026  13.338  < 2e-16 ***
## physical.scores  1.456113   0.123001  11.838  < 2e-16 ***
## food.insecure    0.056009   0.022993   2.436   0.0149 *
## food.limited    -0.039100   0.008537  -4.580 4.82e-06 ***
## income.log      -3.740556   0.486821  -7.684 2.06e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.868 on 3111 degrees of freedom
## Multiple R-squared:  0.2679, Adjusted R-squared:  0.2669
## F-statistic: 284.5 on 4 and 3111 DF,  p-value: < 2.2e-16
```

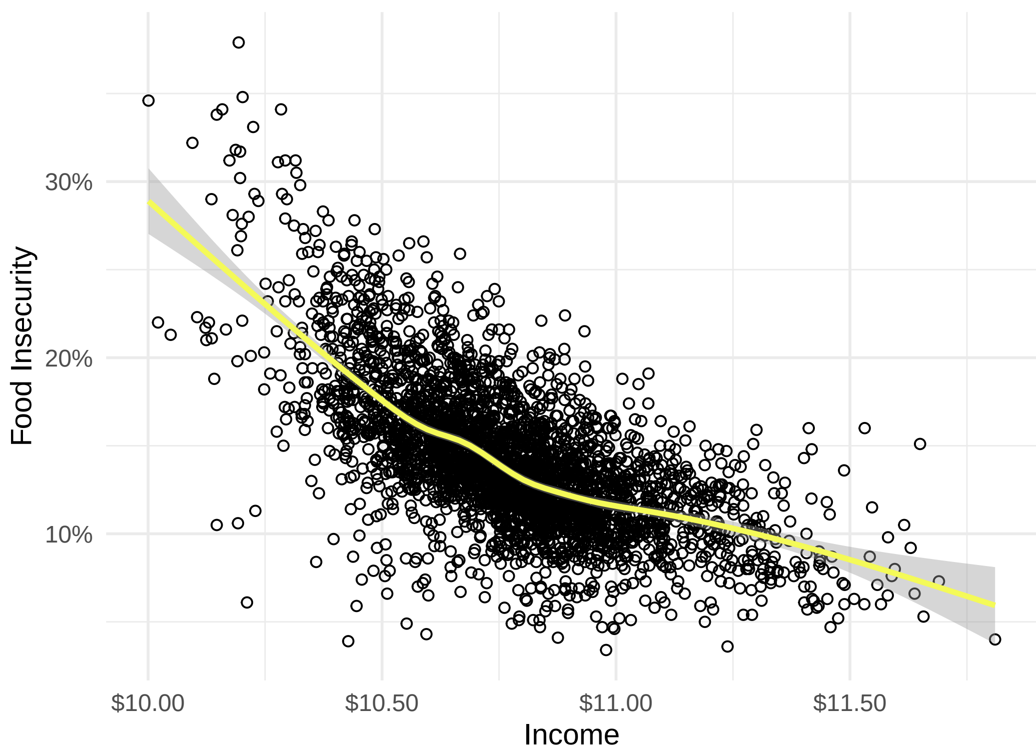
Looking at the summary of the linear regression model above, P-value of all the factors are smaller than 0.05, meaning that they are statistically significant. The fit model has 0.2679 of R-squared value, which is quite low, but it does not always mean inherently bad. Regardless of low R-squared value, I can still draw the fact that the independent variable is associated with changes of the dependent variables in terms of high P-value, as I discussed. Especially, because this study aims to inspect what factors influence on the adult obesity rate, I could tell that this model is applicable to support proving my idea and hypotheses.

The more details of each attribute will be stated here. Above all, the physical scores, which is the latent trait of physical data, seems statistically significant the most among the four candidates. A county that has a higher physical score is likely to have a higher adult obesity prevalence rate. Next, food insecure shows the positive relationship to adult obesity, and this indicates respondents who have insufficient opportunity to healthy food are more likely to be overweight at the county level. However, as its P-value is relatively higher than other factors, it may have a weaker impact to estimate the obesity prevalence rate. In addition, the percentage of people who have limited access to healthy foods show a negative relationship with obesity, which miss my expectation. I assume it's because the samples are extremely biased, so they are not good enough to be examined and this data has some outliers that disturb to conduct a better experiment. Lastly, the county that has a lower median income are apt to have higher adult obesity rate, and this result underpins my hypothesis of the relationship between obesity prevalence rate and income. The one thing I additionally found is that physical inactiveness & food insecurity and income are related as shown in a graph below. This proves the idea that low income is associated with low levels of physical activity and limited choices of food.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

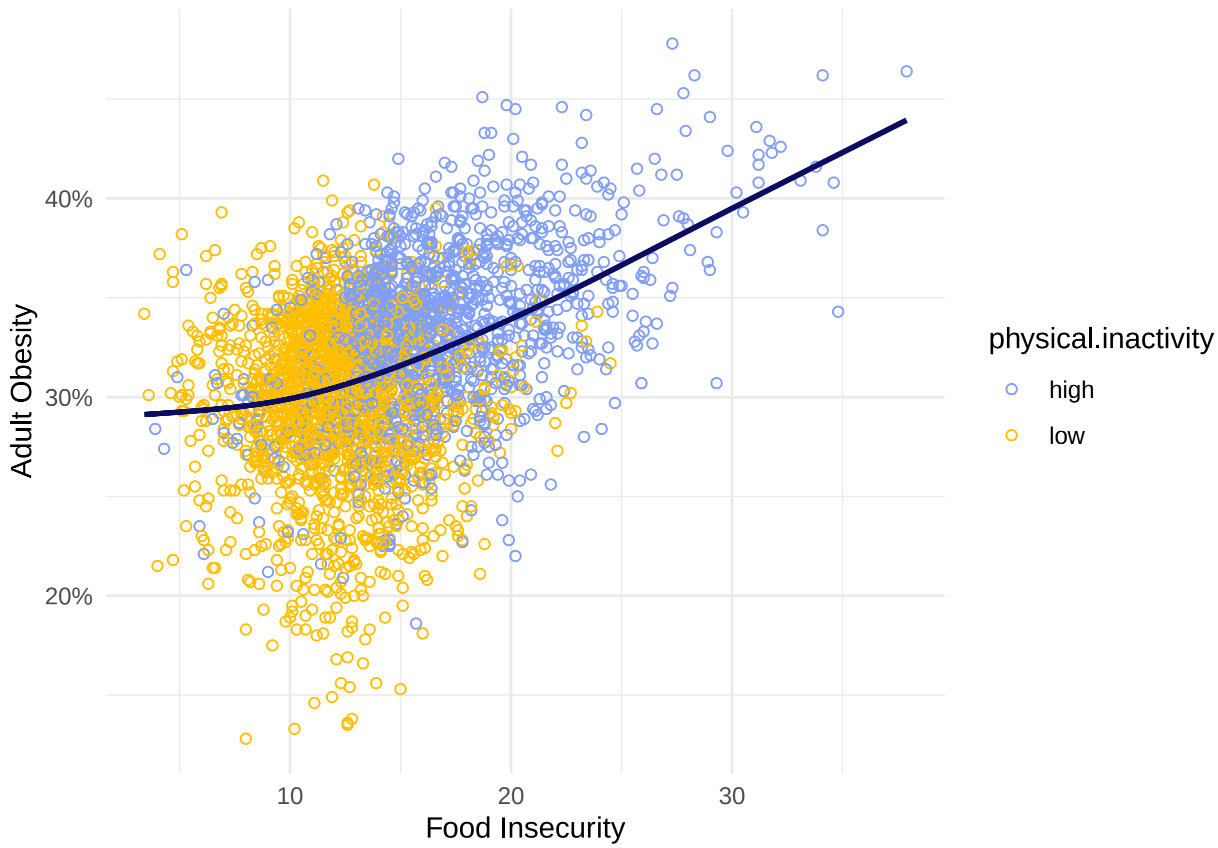



```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Moreover, as a graph below illustrates, I found out that food insecurity and physical inactivity are also correlated. Physical inactivity group (low and high) was decided using the latent trait of physical data. In other words, people who more experience food insecurity are likely to undergo physical inactiveness, and this possibly lead them to be overweight.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Conclusion

A number of approaches to discover the leading predictors of adult obesity have been carried out in this study. First of all, physical inactivity, accessibility to exercise facilities, chronic physical issues, and frequent physically unhealthy day have been verified as more prominent factors of adult obesity, compared to the obtainability of healthy food. However, accessibility to exercise had a relatively lower impact than other predictors do, and it might have been caused by data skewness. Moreover, food insecurity was found more critical for obesity rate, whereas limited access to healthy food was comparatively not. This might be a consequence of some possible limitations of data collection or its reliability. I anticipated that limited access to fresh products would result in higher obesity rate, but this information was not a key contributor to the increasing obesity rate, shown by the experiment. However, median household income showed a strong relationship with obesity prevalence. I also came up with the empirical correlation between income and physical inactiveness & food insecurity, as some papers have theoretically predicted.

Bibliography

[1] Ploeg, M. (2012). Access to affordable and nutritious food-measuring and understanding food (pp. 51-59). Bibliogov.

[2] Moore, L., Diez Roux, A., Nettleton, J., & Jacobs, D. (2008). Associations of the Local Food Environment with Diet Quality–A Comparison of Assessments based on Surveys and Geographic Information Systems: The Multi-Ethnic Study of Atherosclerosis.*American Journal Of Epidemiology,*167(8), 917-924. https://doi.org/10.1093/aje/kwm394

[3] French, S., Tangney, C., Crane, M., Wang, Y., & Appelhans, B. (2019). Nutrition quality of food purchases varies by household income: the SHoPPER study.*BMC Public Health,*19(1). https://doi.org/10.1186/s12889-019-6546-2

[4] Michimi, A., & Wimberly, M. (2010). Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States.*International Journal Of Health Geographics,* 9(1), 49. https://doi.org/10.1186/1476-072x-9-49

[5] Warburton, D. (2006). Health benefits of physical activity: the evidence.*Canadian Medical Association Journal*,174(6), 801-809. https://doi.org/10.1503/cmaj.051351

[6] Giovannucci, E., Colditz, G. A., Stampfer, M. J., & Willett, W. C. (1996). Physical activity, obesity, and risk of colorectal adenoma in women (United States).*Cancer causes & control : CCC,*7(2), 253–263. https://doi.org/10.1007/BF00051301

[7] French, S., Tangney, C., Crane, M., Wang, Y., & Appelhans, B. (2019). Nutrition quality of food purchases varies by household income: the SHoPPER study.*BMC Public Health,*19(1). https://doi.org/10.1186/s12889-019-6546-2

[8] *About Rural Health* | CSELS | OPHSS | CDC. Cdc.gov. (2017). Retrieved 9 June 2020, from https://www.cdc.gov/ruralhealth/about.html.

[9] Jebb, S., Kopelman, P., Butland, B., McPherson, K., Thomas, S., & Mardell, J. (2007). Tackling Obesities: Future Choices - Project Report.*Obesity Reviews*, 8(s1), vi-ix. https://doi.org/10.1111/j.1467-789x.2007.00344.x

[10] Shuval, K., Li, Q., Gabriel, K., & Tchernis, R. (2017). Income, physical activity, sedentary behavior, and the 'weekend warrior' among U.S. adults.*Preventive Medicine,*103, 91-97. https://doi.org/10.1016/j.ypmed.2017.07.033

[11] https://www.countyhealthrankings.org/

[12] https://www.cdc.gov/