Exploring the affect of physical activities to the obesity of US adults: An analysis of NHANES {aplore3} Data

Course Code : MAT 5317

group members : Ann Warnakulasuriya - awarn054@uottawa.ca Lupa Deb - ldeb078@uottawa.ca

Introduction

The National Health and Nutrition Examination Survey (NHANES) ("The Third National Health and Nutrition Examination Survey (NHANES III, 1988-94) Reference Manuals and Reports-Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94-Analytic and Reporting Guidelines-Weighting and Estimation Methodology-Accounting for Item Nonresponse Bias-Field Operations Reference Manuals (by Topic) Next Page" 1996) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey combines interviews and physical examinations. NHANES is a major program of the National center for health statistics (NCHS). The NHANES program began in the early 1960s and has been conducted as a series of surveys focusing on different population groups or health topics. The survey became a continuous program in 1999. It basically examines a nationally representative sample of about 5,000 persons each year. NHANES findings are the basis for national standards for such measurements as height, weight and blood pressure.

Background

Obesity has become one of the most important public health problems worldwide, which suggests the need for evidence-based dietary strategies for weight loss and it's maintenance. More than 650 million adults worldwide suffer from obesity and the prevalence of this condition has increased rapidly during the past 50 years. The world Health Organization (WHO) defines over-weight, and obesity as abnormal or excessive fat accumulation that presents a risk to health (WHO,2016a). A body mass index (BMI) >= 25 kg/m2 is generally considered overweight, while obesity is considered to be a BMI >= 30 kg/m2. Obesity is strongly associated with type 2 diabetes mellitus(T2DM); cardiovascular diseases including myocardial infarction and stroke; osteoarthritis; obstructive sleep apnea; depression and some types of cancer, such as breast, ovarian, prostate, liver, kidney and colon cancer.Therefore this study was conducted to investigate the behaviour of obesity among US adults

Research Questions

1) Studying the effect from Physical activities to Obesity.
2) Studying the effect from Cholesterol and Blood Pressure to the Obesity.

Study Design and Variables

US Department of Health and Human Services has mentioned in their National Heath and Nutrition Examination Survey:Plan and Operations,1999 - 2010 article (Health Statistics 2020) how they have conducted the survey, methods of interviewing, methods of sampling etc. in order to collect the data related to vital and health statistics. When the sample design is studying, a complex, multistage design has been used to select a samples. First they have selected samples at each stage in order to have a more representative sample. First they selected Primary sampling units then within each PSUs again selected clusters, even they have samples within each households selected. They have also assigned

sample weights to each person in the sample as a measure of number of people represented by that particular sample person. Different types of questionnaires has been used to collect the data such as Relationship questionnaire, Sample person questionnaire and Family questionnaire.They have used Mobile Examination centers to get the information regarding the health and nutrition such as hearing, weight, etc.

The data the NHANES (Health Statistics 1999) gathered used by them to give the insights of National Health and Nutrition level. National center for health and statistics has published a report on study they conducted using the NHANES data and got the insights regarding the Obesity.

Zhilei Shan and Frank B. Hu (Shan et al. 2019) has published a paper regarding the trends in dietart Carbohydrates, Proteins and Fat intakes and Diet quality among US adults using the NHANES data. They have used 9 survey cycles from 1999 to 2016 to find the trends. They have used different statistical methods such as clustering, stratification, etc. in order to obtain sample weights.

Data Description

The NHANES {aplore3} Data has 6482 observations and 21 variables. The 21 variables of the dataset is as follows;

Characteristics	Description
Id	Identification code (1-6482)
gender	Gender (1: Male, 2: Female)
age	Age at Screening (Years)
marstat	Marital Status (1: Married, 2: Widowed, 3: Divorced, 4: Separated, 5: Never Married, 6: Living Together)
samplewt	Statistical Weight
PSU	Primary sampling units (1,2)
Strata	Statified random sampling (1-15)
tchol	Total cholesterol (mg/dL)
hdl	HDL- cholesterol (mg/dL)
sysbp	Systolic blood pressure (mmHg)
dbp	Diastolic blood pressure (mmHg)
wt	Weight (kg)
ht	Standing Height (cm)
bmi	Body mass index (kg/m2)
vigwrk	Vigorous work activity (1: Yes, 2: No)
modwrk	Moderate work activity (1: Yes, 2: No)
wlkbik	Walk or Bicycle (1: Yes, 2: No)
vigrecexr	Vigorous Recreational activities (1: Yes, 2: No)
modrecexr	Moderate Recreational activities (1: Yes, 2: No)
sedmin	Minutes of Sedentary activity per Week (1: Yes, 2: No)
obese	BMI>35 (1: No, 2: Yes)

Table 1 : Data Description

Data Composition

According to the Table 1 which showed the description of the variables there are 21 variables were studies in this study. The composition of these variables in the dataset is presented in this section.

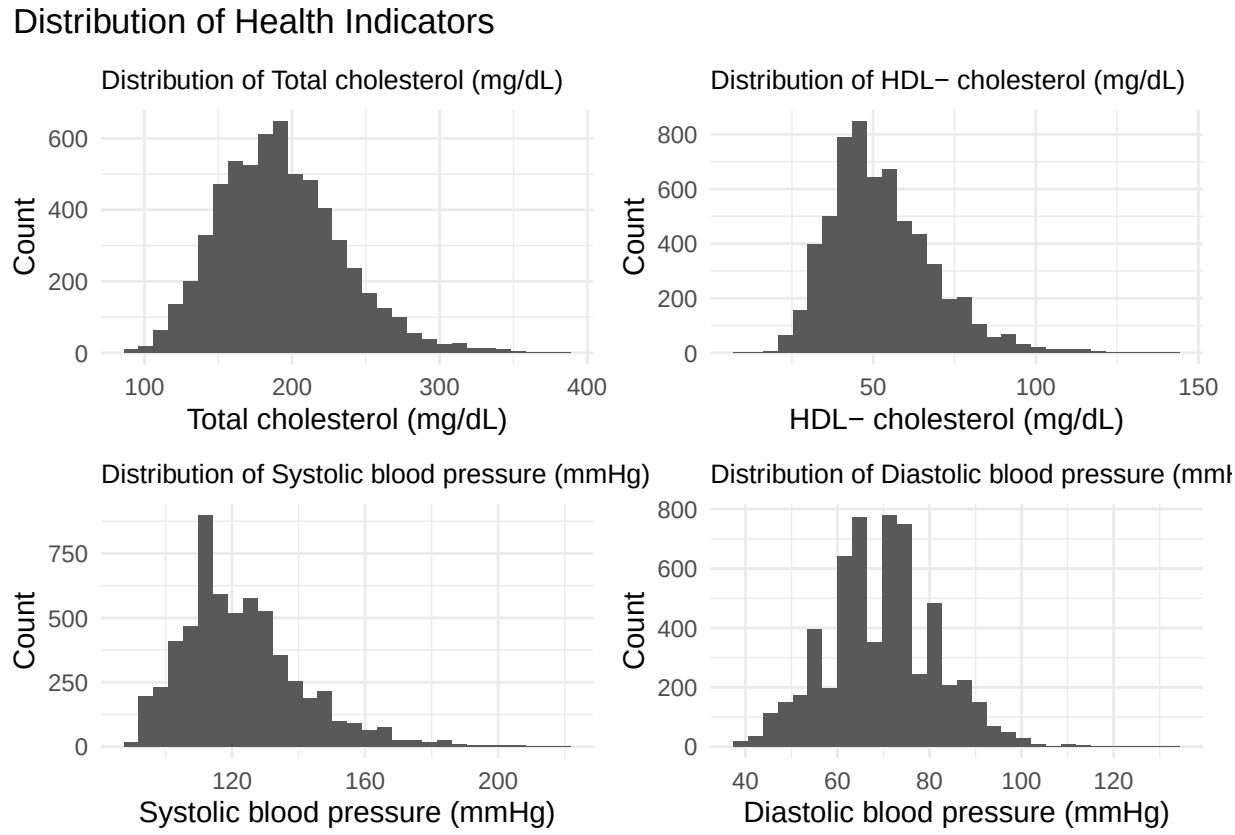
Table 2 shows the data composition of the nhanes data set after removing the missing data in the "Obese" variable which was 37 observations. Table 2 shows that there are almost equal composition in gender, Primary sampling units (PSU) and Stratified random sampling unit (strata). Therefore it can be assumed that the data has been sampled from the population in a much representative manner. There are two primary sampling units and 15 strata in the data.

Data Frame Summary
DF_1
Dimensions: 6445 x 5
Duplicates: 2079

Variable	Stats / Values	Freqs (% of Valid)	Graph
gender\	1\. Male\	3144 (48.8%)\	IIIIIIIII \
[factor]	2\. Female	3301 (51.2%)	IIIIIIIIII
age\	Mean (sd) : 46.4 (19.4)\	65 distinct values	:\
[integer]	min < med < max:\		: \ \ . . : \ \ \ \ . \ \ :\
	16 < 46 < 80\		: : : : : : : : . :\
	IQR (CV) : 32 (0.4)		: : : : : : : : : :\
			: : : : : : : : : :
marstat\	1\. Married\	3004 (51.6%)\	IIIIIIIIII \
[factor]	2\. Widowed\	504 (8.7%)\	I \
	3\. Divorced\	640 (11.0%)\	II \
	4\. Separated\	193 (3.3%)\	\
	5\. Never Married\	1023 (17.6%)\	III \
	6\. Living Together	454 (7.8%)	I
psu\	Min : 1\	1 : 3157 (49.0%)\	IIIIIIIII \
[integer]	Mean : 1.5\	2 : 3288 (51.0%)	IIIIIIIIII
	Max : 2		
strata\	1\. 1\	497 (7.7%)\	I \
[factor]	2\. 2\	507 (7.9%)\	I \
	3\. 3\	553 (8.6%)\	I \
	4\. 4\	538 (8.3%)\	I \
	5\. 5\	431 (6.7%)\	I \
	6\. 6\	497 (7.7%)\	I \
	7\. 7\	468 (7.3%)\	I \
	8\. 8\	433 (6.7%)\	I \
	9\. 9\	421 (6.5%)\	I \
	10\. 10\	422 (6.5%)\	I \
	[5 others]	1678 (26.0%)	IIIII

Table 2 : Data Composition

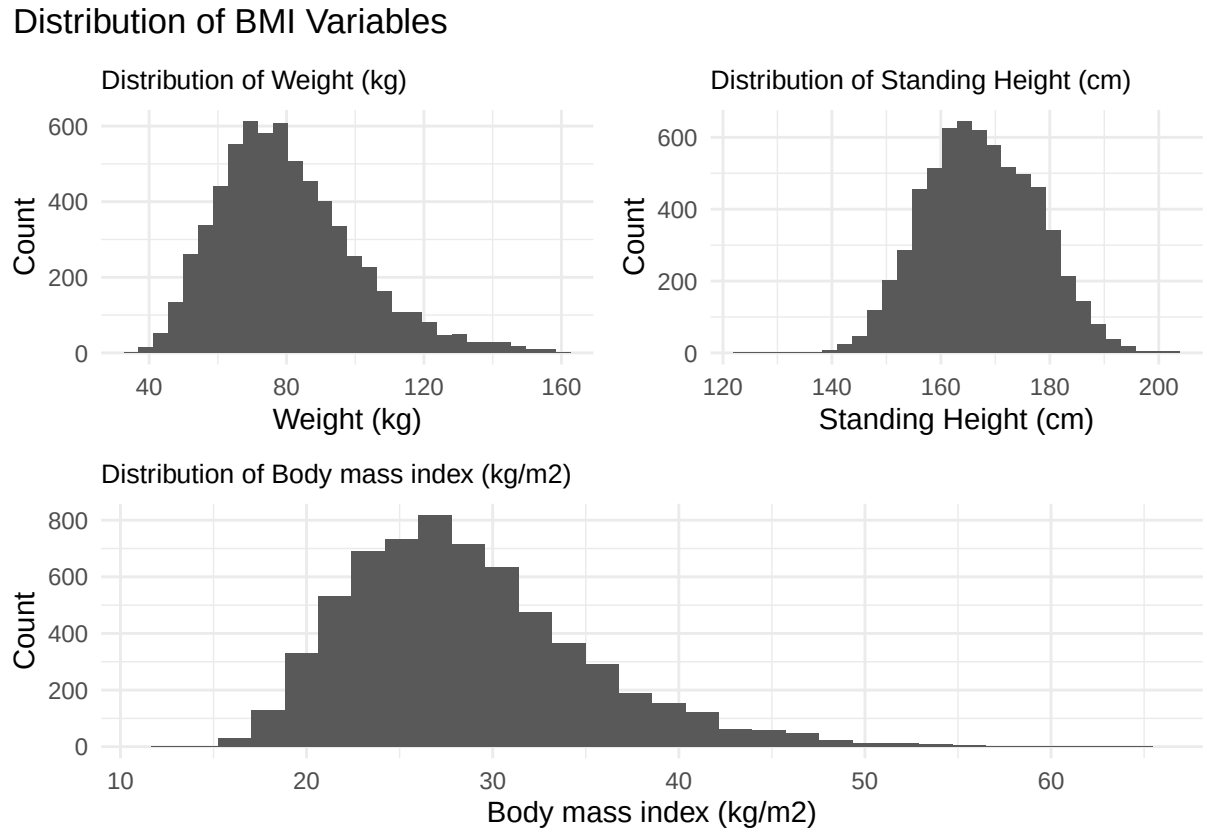
The nhanes data set consists of both categorical and continuous data. The health indicators such as cholesterol level and blood glucose level has shown in continuous form where as normally it is considered in interval scale.



>Figure 1 : Distribution of Cholesterol Variable

Figure 1 shows that Cholesterol variables show approximately a symmetric distributions. Therefore in order to get the relationship between the weight and the Cholesterol levels and also among the cholesterol level and the activity level of a person it has categorized and analysed in the next section.

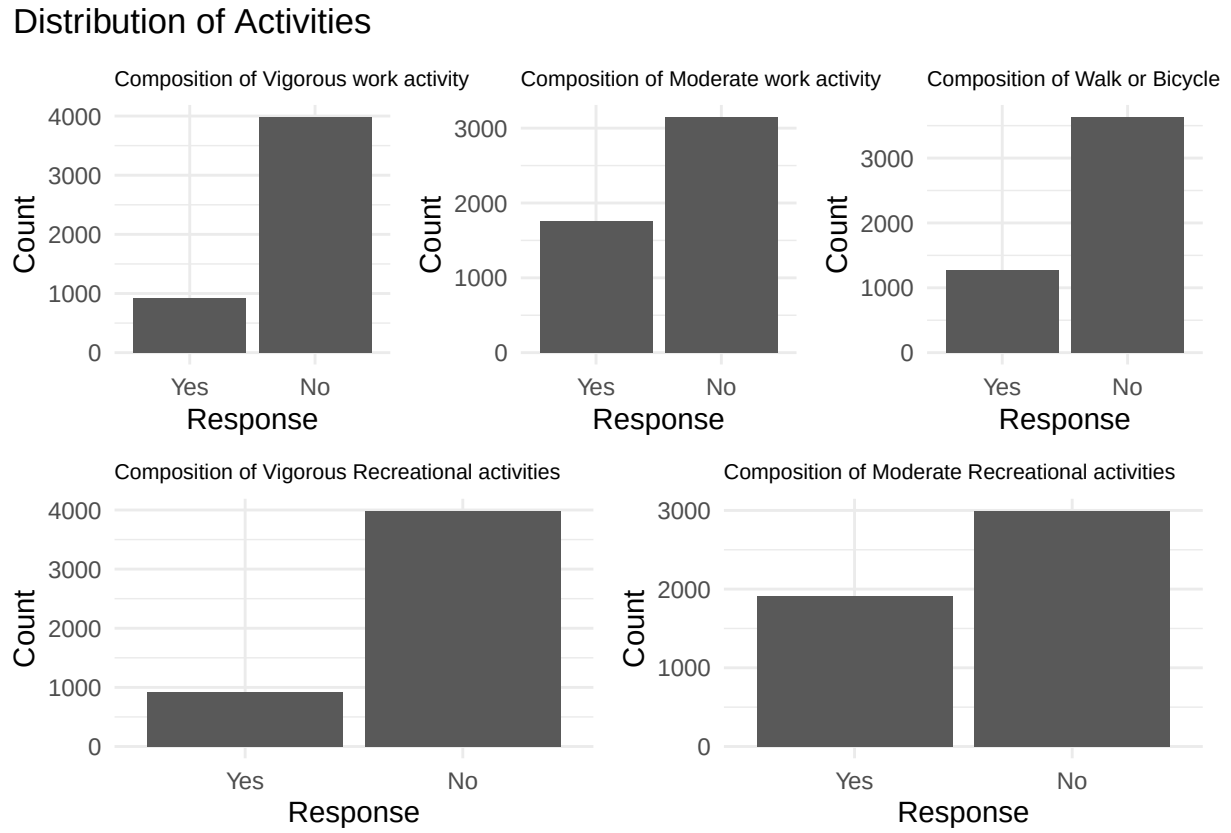
In order to calculate the obesity of a person the BMI has be to calculate with respect to the Weight and the height of a person. The distribution of those variables has shown below.



>Figure 2 : Distribution of BMI Variables

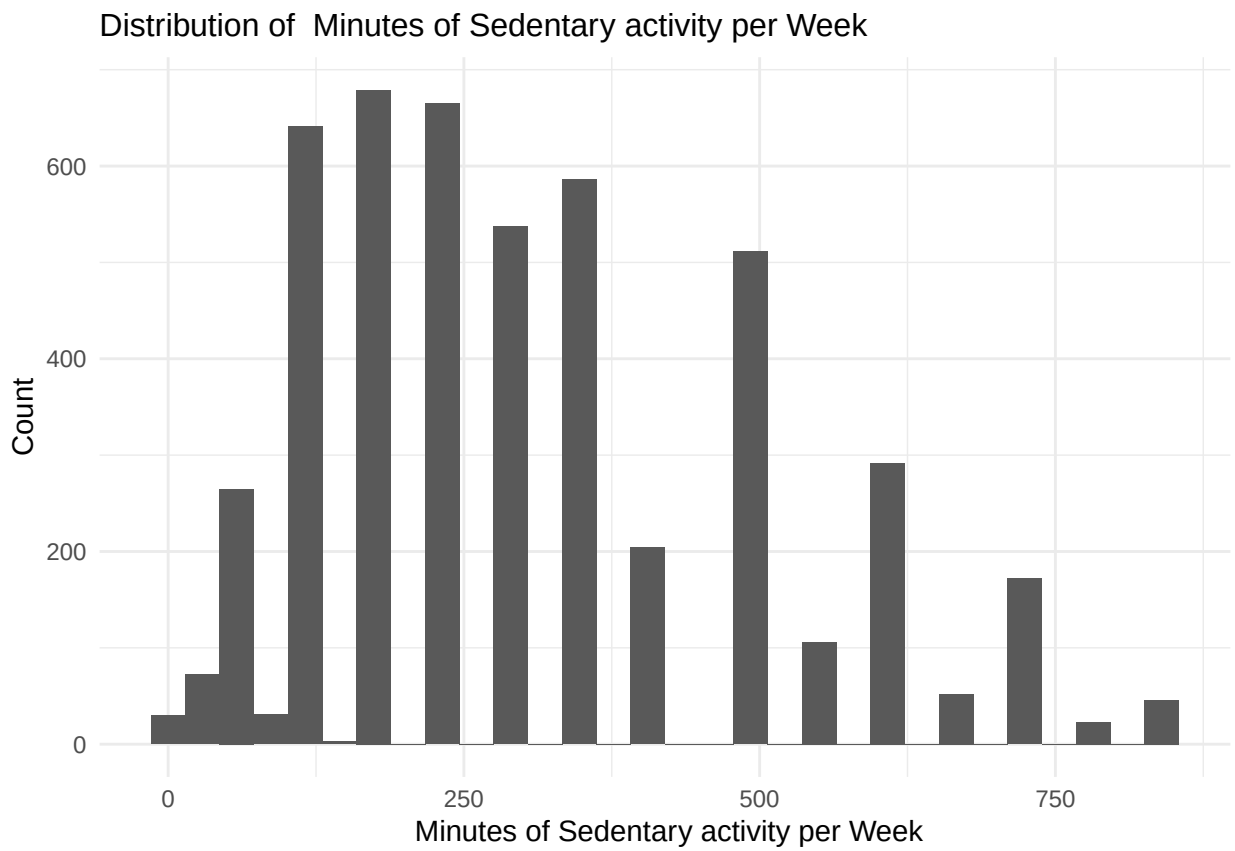
It can be identified from the figure 2 that the weight and height are normally distributed with the BMI. Since the BMI is from the weight and the height, these variables was not analysed further.

The main objective of the study was the activity level of a US person. Below plot shows percentages of involvement of US people in activities.



>Figure 3 : Distribution of Activities

According to the above plot most of the US people are not involved in significant activities. Whereas the propotion of people who walk or bicycle is less than 1/3 of the population.

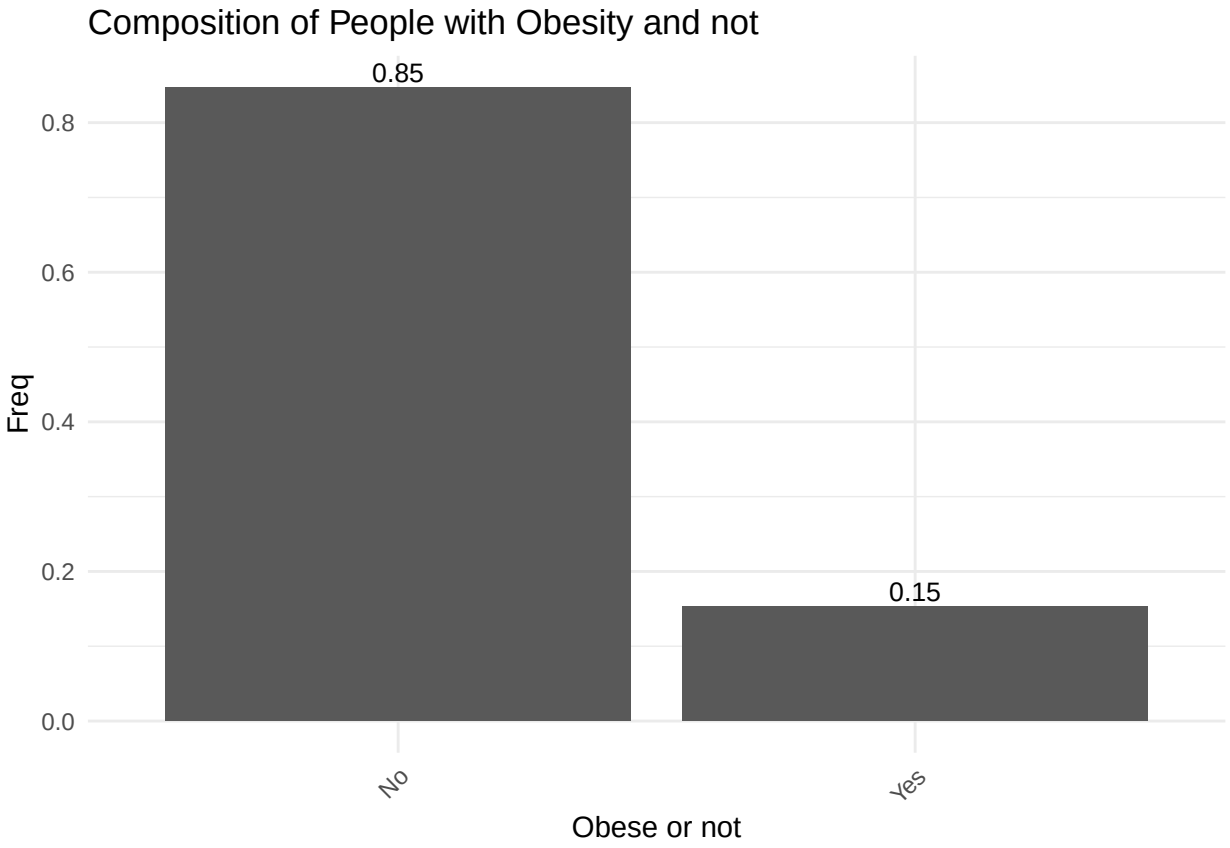


>Figure 4 : Distribution of Activities

According to the Figure 3 more than 50% of the people has spent less than 250 min (< 5 hrs) per week.

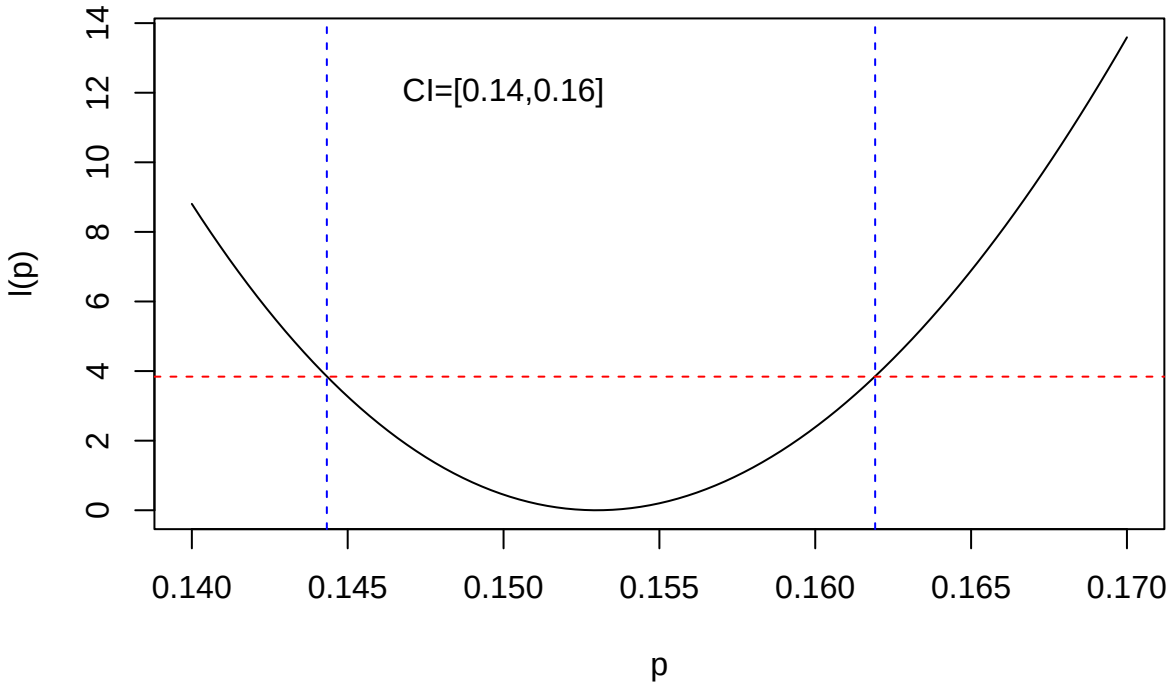
Data Analysis

Proportion of Obesity



>Figure 5 : Composition of People with Obesity and not

Fig 5 shows that there is only 15% has been categorized as people with obesity among the US population. The probability that a person will be a suffers with obesity or not is a binomial distribution therefore the parameters for the distribution was obtained as below in a Confidence Interval format.



>Figure 6: Likelihood Ratio CI for proportion obesity at 95% Confidence

There it can be concluded that the proportion of people with obesity is between 14% to 16% at 95% confidence according to the data. The Below table which shows the different Confidence Interval types also shows that the proportion of people with obesity is between approximately 14 % to 16%.

	method	x	n	mean	lower	upper
1	agresti-coull	986	6445	0.1529868	0.1444029	0.1619841
2	asymptotic	986	6445	0.1529868	0.1441984	0.1617752
3	bayes	986	6445	0.1530406	0.1442844	0.1618573
4	cloglog	986	6445	0.1529868	0.1443179	0.1618916
5	exact	986	6445	0.1529868	0.1442793	0.1620095
6	logit	986	6445	0.1529868	0.1444037	0.1619835
7	probit	986	6445	0.1529868	0.1443656	0.1619426
8	profile	986	6445	0.1529868	0.1443371	0.1619123
9	lrt	986	6445	0.1529868	0.1443375	0.1619262
10	prop.test	986	6445	0.1529868	0.1443296	0.1620611
11	wilson	986	6445	0.1529868	0.1444053	0.1619817

Distribution of Obesity with respect to the Age and the Gender.

In order to find whether there is any relationship between the Obesity, Age and Gender, chi squared statistics was used as follows;

	obese	
gender	Yes	No
Female	0.1875189	0.8124811
Male	0.1167303	0.8832697

Table 3: Proportion of Obesity across Gender

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  .
X-squared = 61.726, df = 1, p-value = 3.947e-15
```

Since p-value is less than 5%, it can be concluded that we have enough evidence to say that there is a relationship between the gender and obesity.

Table 3 above shows the proportion of males and females suffers with obesity. Below information shows that obesity is approximately 1.61 times more likely to occur in females compared to males.

```
        obese
gender    Yes   No
  Female  619 2682
  Male    367 2777

rel. risk    lwr.ci    upr.ci
 1.606429  1.425559  1.810248
```

Table 4 : Relative Risk of Obesity between gender

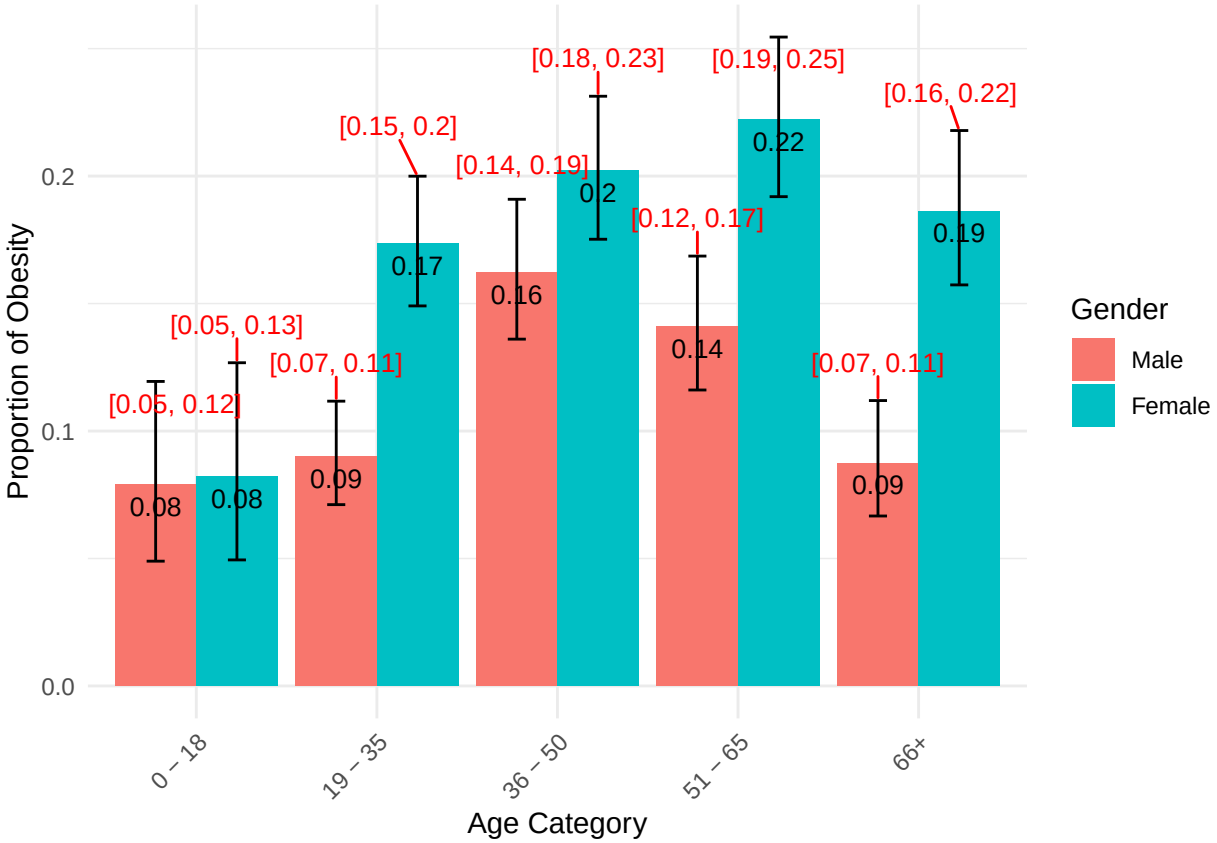
```
            obese
Age_Category        Yes         No
     66+     0.13771997 0.86228003
     51 - 65 0.18136909 0.81863091
     36 - 50 0.18334409 0.81665591
     19 - 35 0.13352941 0.86647059
     0 - 18  0.08050847 0.91949153
```

Table 5: Proportion of Obesity across Age

```
    Pearson's Chi-squared test

data:  .
X-squared = 46.277, df = 4, p-value = 2.157e-09
```

Therefore it can be concluded that both gender and age has an association for the Obesity. As a result the proportion of Obesity in each gender wise and age-wise was obtained as below using the Clopper Pearson Confidence Interval



>Figure 7: Distribution of Confidence interval of proportion of Obesity across Age and Gender.

From the above plot it can be identified that females of age group 51 to 65 years suffers most from obesity while males of the age group 36 years to 50 years suffers from obesity than other age groups. However overall females has higher probability of suffering from Obesity than that of males.

Effect from Physical activities to Obesity

Mainly there are 5 Physical activities have considered with a Behavior activity which have an effect to the physical activity. The 5 main physical activities considered here are whether the person involved in, 1) Vigorous Work 2) Moderate Work 3) Walking or Biking 4) Vigorous Exercise 5) Moderate Exercise

The contrast of physical activities which is "How much time a person spend in Sedentary activity". Depend on the Sedentary activity it might help in mental health but not the physical health according to studies.

```
      obese
vigwrk  Yes   No
   No   820 4495
   Yes  166  963
```

```
rel. risk    lwr.ci    upr.ci
 1.049292  0.899785  1.223641
```

Table 6 : Relative Risk of Obesity between Vigorous Work

Table 6 shows that obesity is approximately 1.05 times more likely to occur in people not involved in vigorous work compared to the ones that are involved in vigorous work.But at 95% confidence there is a possibility of having less risk of obesity for the people not involved in vigorous activity than that of people not involved in vigorous activity.

```
      obese
modwrk  Yes   No
   No   667 3597
   Yes  319 1861
```

```
rel. risk    lwr.ci    upr.ci
1.0689920 0.9453234 1.2088390
```

Table 7 : Relative Risk of Obesity between Moderate Work

Table 7 shows that obesity is approximately 1.07 times more likely to occur in people not involved in moderate work compared to the ones that are involved in moderate work. But in 95% confidence there are possibilities that this may may happen vise-versa since lower CI is 0.95. Which is there are possibilities that obesity is approximately 0.99 times more likely to occur in people not involved in moderate work compared to the ones that are involved in moderate work.

```
      obese
wlkbik  Yes   No
   No   780 3864
   Yes  206 1594
```

```
rel. risk    lwr.ci    upr.ci
 1.467600  1.271541  1.693889
```

Table 8 : Relative Risk of Obesity between Walking and Biking

Table 8 shows that obesity is approximately 1.47 times more likely to occur in people not involved in Walking or Biking compared to the ones that are involved in Walking or Biking.It is 95% Confidence that the True RR lie between 1.27 and 1.69. Therefore it can be concluded that we are 95% confidence that the risk of having obesity for people not involved in waling or biking is more than 27% but less than 70% than of the people involved in walking or biking.

```
         obese
vigrecexr  Yes   No
      No   866 4179
      Yes  120 1279
```

```
rel. risk    lwr.ci    upr.ci
 2.001212  1.669678  2.398577
```

Table 9 : Relative Risk of Obesity between Vigorous Exercise

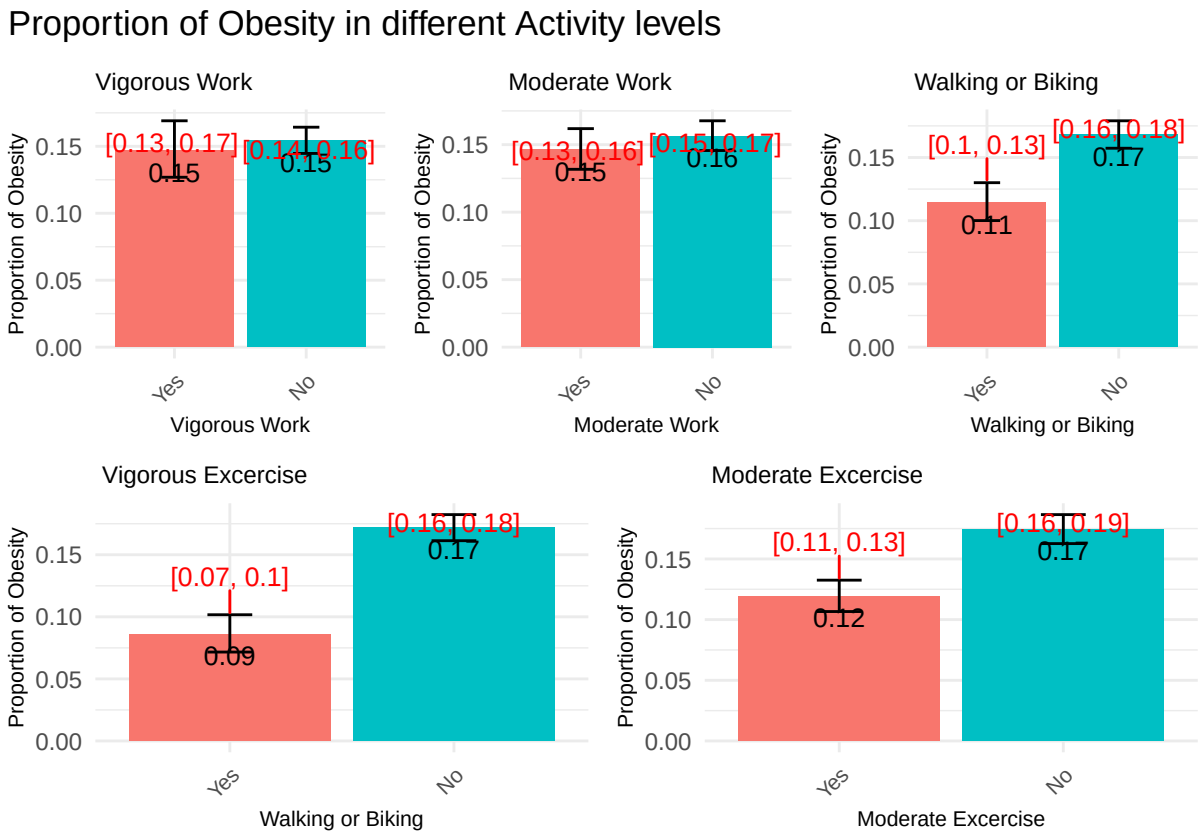
Table 9 shows that obesity is approximately 2 times more likely to occur in people not involved in Vigorous Exercises compared to the ones that are involved in Vigorous Exercise. We are 95% Confidence that the risk of having obesity if a person not doing any Vigorous Exercises is more than 100% that of a one doing Vigorous Exercises.

```
         obese
modrecexr  Yes   No
      No   690 3268
      Yes  296 2189
```

```
rel. risk    lwr.ci    upr.ci
 1.463551  1.289640  1.660915
```

Table 9 : Relative Risk of Obesity between Moderate Exercise

Table 9 shows that obesity is approximately 1.46 times more likely to occur in people not involved in Moderate Exercises compared to the ones that are involved in Moderate Exercise. We are 95% Confidence that the risk of having obesity if a person not doing any Moderate Exercises is more than 29% but less than 66% that of a one doing Moderate Exercises.



>Figure 8: Proportion of Obesity in different Activity levels

From the above figure and analysis it can be identified that when people involved in some kind of exercise like biking, walking, vigorous or moderate the probability of being obese is lower than not doing it. According to the above plot doing vigorous exercises affect a lot to the obesity where the probability is lower than other activities.

In order to find the relationship between the Sedentary activity and the Obesity, minutes of Sedentary activity per Week was categorised and used chi-squared test identify the assocation between those variables.

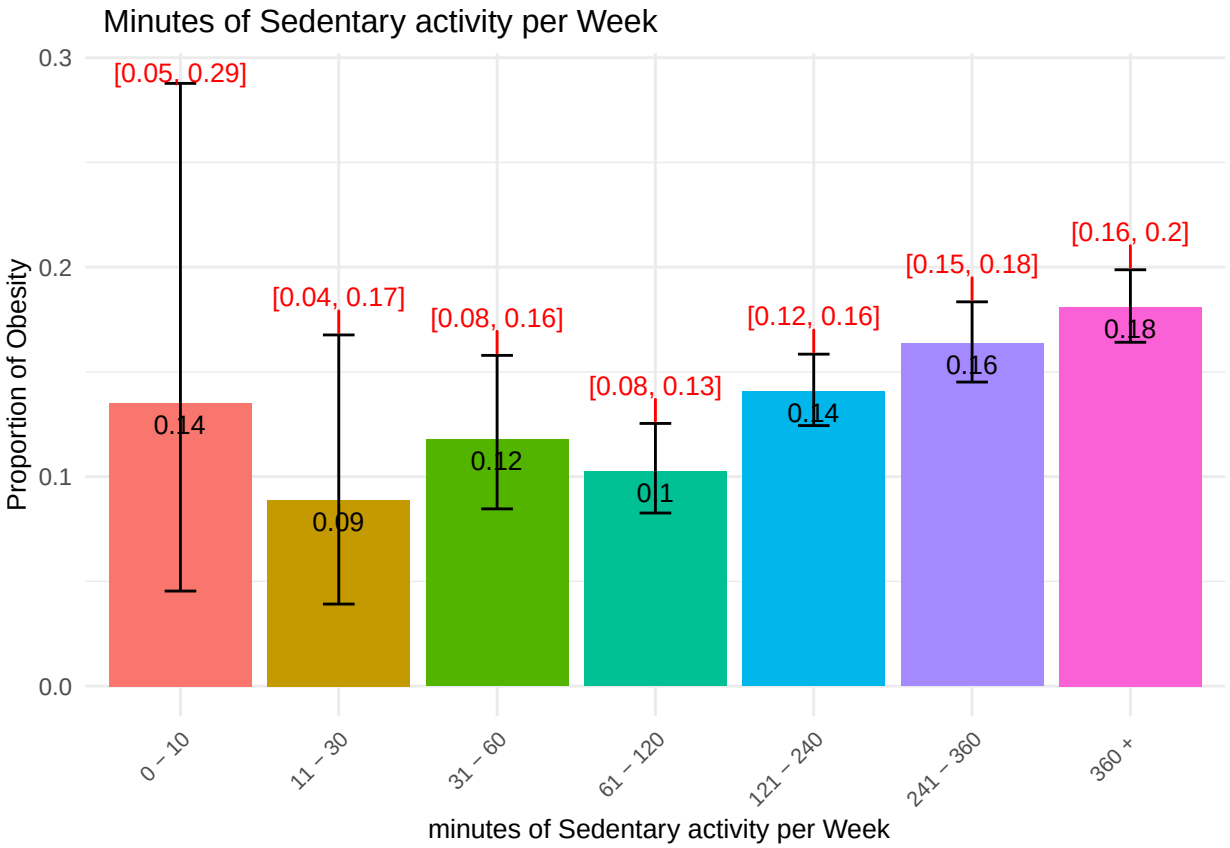
```
                obesity
Sedmin_Category         Yes          No
      360 +     0.055616654 0.251688924
      241 - 360 0.038177533 0.195129615
      121 - 240 0.036606441 0.223409269
      61 - 120  0.013197172 0.115475255
      31 - 60   0.005970149 0.044776119
      11 - 30   0.001256874 0.012882954
      0 - 10    0.000785546 0.005027494


    Pearson's Chi-squared test

data:  .
X-squared = 37.34, df = 6, p-value = 1.511e-06
```

Table 10 : Proportion table of Obesity across minutes of Sedentary activity per Week

Since the p-value is less than 0.05 it can be concluded that there is an association between the Sedentary activity and the Obesity



>Figure 9: Minutes of Sedentary activity per Week with obesity Proportion

Above figure shows that there is a slight increase in obesity percentage when the time per sedentary increases. Therefore it can be concluded that the more the people been idle the more risk to get obesity.

Effect from Health Indicators to Obesity

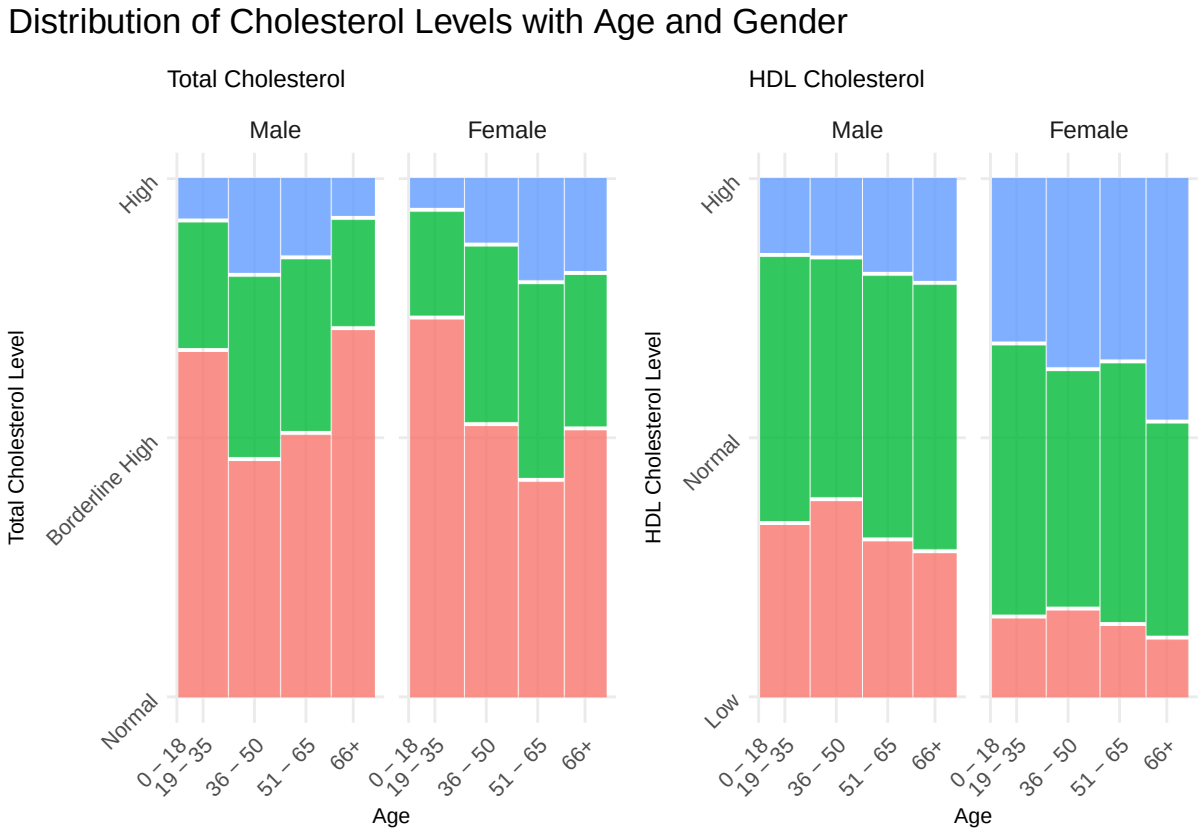
There are four indicators used in this study to measure the health other than BMI which are Total Cholesterol level, HDL cholesterol level, Systolic blood pressure and Diastolic blood pressure. In figure 1 it was observed that all those are distributed in a approximately normal distribution. Since these indicators are studied in a interval scale to find effect of those indicators to study the health conditions, those variable were categorized as below in a given threshold.

1. **Total Cholesterol (mg/dL)**:
   - Normal: Less than 200
   - Borderline High: 200-239
   - High: 240 and above
2. **HDL Cholesterol (mg/dL)**:
   - Low (for men): Less than 40
   - Low (for women): Less than 50
   - Normal: 40-59
   - High: 60 and above
3. **Systolic Blood Pressure (mmHg)**:
   - Normal: Less than 120
   - Elevated: 120-129
   - Hypertension Stage 1: 130-139
   - Hypertension Stage 2: 140 and above
4. **Diastolic Blood Pressure (mmHg)**:

- Normal: Less than 80
- Elevated: 80-89
- Hypertension Stage 1: 90-99
- Hypertension Stage 2: 100 and above

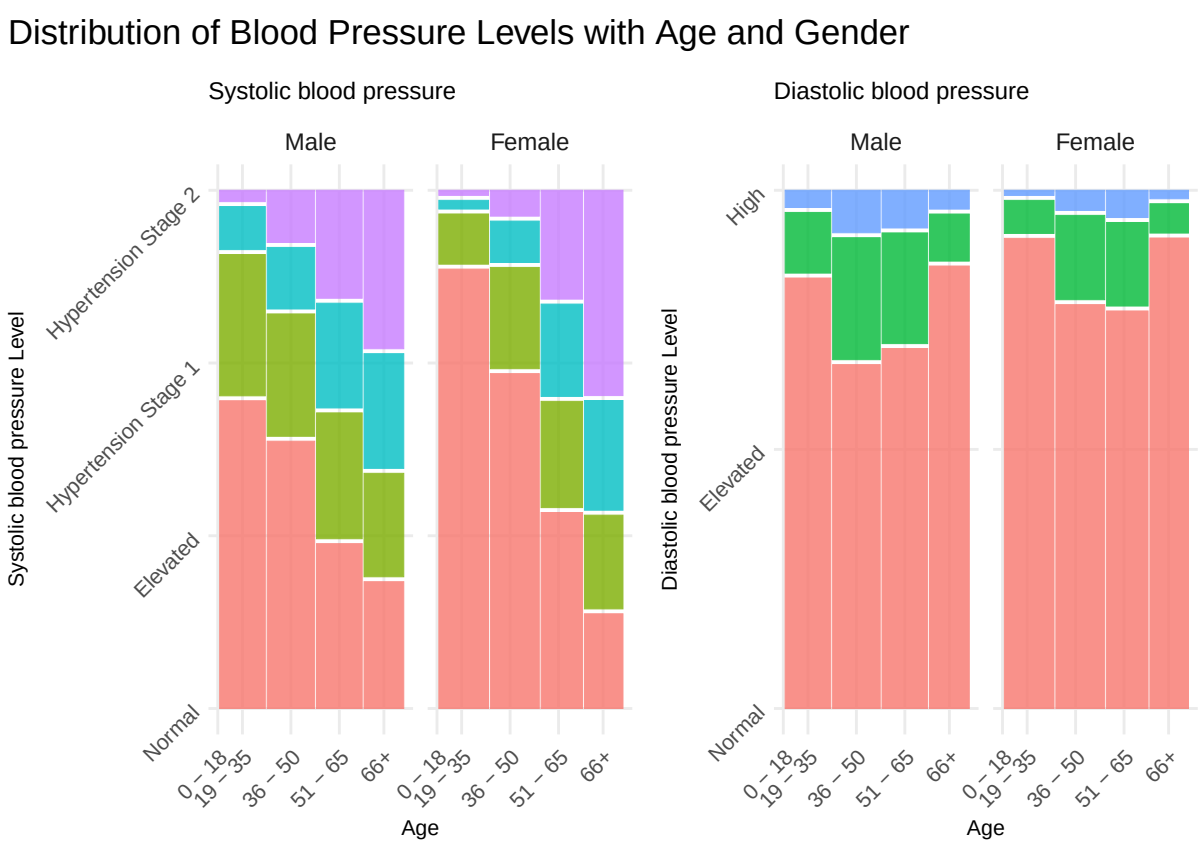
The above thresholds were obtianed from "https://www.medicalnewstoday.com/articles/315900" and "https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/blood-pressure/art-20050982"

The effect of Cholesterol and Blood Pressure Levels has analysed below. In order to find the composition of cholesterol levels with respect to the age and the gender mosaic plots were as below.



>Figure 10: Distribution of Cholesterol Levels with Age and Gender

It can be identified that females have higher proportion of HDL Cholesterol levels when compared with males in each age group, where as its been slightly increasing with the age. When considering the Total cholesterol levels Figure 10 cleary state that males between age 19 to 65 has higher amount of Cholesterol than other age groups. Even the same behavior is observed in the Female age groups.



>Figure 11 : Distribution of Blood Pressure Levels with Age and Gender

Figure 11 state that in both males and females the systolic blood pressure is drastically increasing with the age, whereas the Diastolic blood pressed has changed much with the age but 66 it has decreased even lower than age 18 group in both males and females.

In order to find whether there is any association between the health indicators and the obesity, chi square test was used and the results has summarized as follows;

	Variable	Chi_Squared	Degrees_of_Freedom	P_Value
1	Total Cholesterol	8.789805	2	1.234008e-02
2	HDL Cholesterol	142.625728	2	1.069589e-31
3	Systolic Blood Pressure	22.258554	3	5.762948e-05
4	Diastolic Blood Pressure	20.149713	2	4.212554e-05

Table 11 : Chi - squared test results between health indicators and the Obesity

Therefore it can be concluded that at 95% confidence that there is an effect from health indicators to the obesity since the p-value is less than 0.05. When we study these relationships further it can identified that

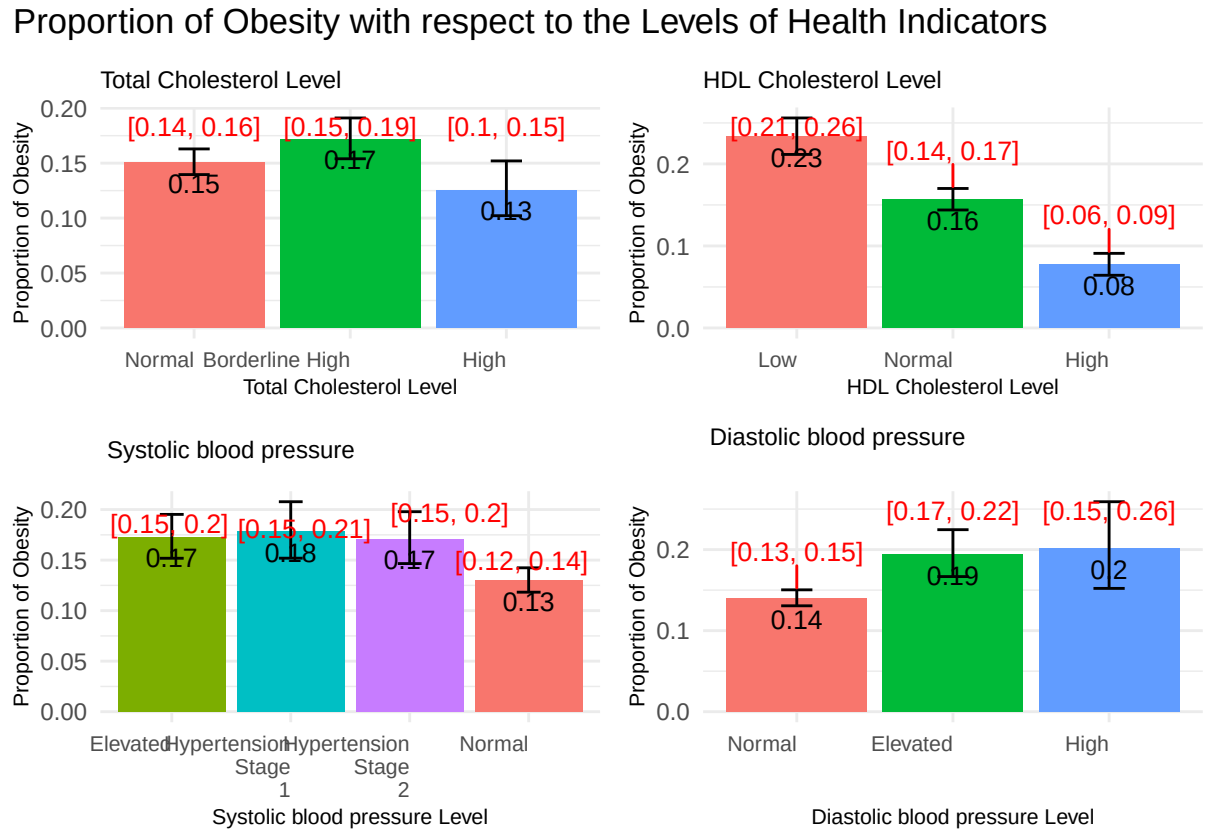


Figure 12 : Proportion of Obesity with respect to the Levels of Health Indicators

The above figure shows that although we assumed that when the cholesterol level in high there is high chance of obesity but data shows that probability of obesity is higher when the Total cholesterol level is Broad-line High only. Where it shows that when the HSL Cholesterol Level increases that the chances of being obese is also getting low.

But it shows that when both systolic and Diastolic blood pressure increases the probability of having obesity also increases.

Conclusion and Discussion

The main Obejective of this study is to find the relationship between Obesity and health/nutrition in nhanes data. Through th background study of this dataset it was found that this is a survey conducted through multiple stages using a complex sampling strategy. But the nahanes datset available in “aplore3” r package has few variables regading the age, gender, marital staus, sample classification, activities, cholesterol levels, blood pressure level and obesity related data.

When exploring the dataset it was observed that the male and female proportion is almost equal and the sampling has also done equally in each psu and srata. Then when exploring the health indicators and BMI related variables it was obvious that all are distributed almost normal since the sample size is more than 6000. From the first sight it was oberved that most of the people which is more than half of the people in US are not into activities such as walking, biking or exercising. But the there is more than 50% of people has spent less than 500 min (~8 hours) in sedernity activities per week.

Although nore than 50% of people doesn’t involved in any significant activities only 15% of population was

shown as suffering in Obesity in nahanes data. Therefore from the statistical test it was concluded that we are 95% confident that the true propotion of Obesity in US population lies between 14% to 16%. This was approved from more than 8 Confidence Intervals generated using different methods such as, agresti-coull, symptotic, etc.

However it was found that females suffer from obesity more than males where as females suffers from obesity mostly in the between 51 to 65 years. where the highest possible percentage is 25%. Where Obesity percentage of females was increased significantlt from age 19 years, whereas males percentage increased and then decreased after the age 50 years. (Fig 7). It was also found that the marital status has no effect to the Obesity.

Then it was observed that all the activity types has an effect to the Obesity. Where it was found that the risk of suffering from Obesity in the absence of vigorous exercise is almost 100% higher than not doing any vigorous exercises. But bot doing moderate or vigorous work sometimes may not be a cause for being obese according to the 95% Confidence intervals.Therefore it can concluded that just because of doing any moderate or vigorous work doesn't guratee not being obese but if any moderate, vigorous is the best and walking or biking has an significant affect to not being obese. It was also proven that when the time taking in sedentary activities increases probability of being obese also increasing.

There were four main health indicators and it was found that females suffer mostly from cholesterol than males even when they get age. Where the systolic blood pressure also behavior the same. Surprising it was oberved that when the HDL cholestrol levels are decreasing then the probability of being obese also decrease. But when the Diastolic blood Pressue increases the percentage of being obese has increased slightly.

References:

Websites :

1. "https://www.medicalnewstoday.com/articles/315900"
2. "https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/blood-pressure/art-20050982"

Health Statistics, National Center for. 1999. "Vital and Health Statistics Report Series 1, Number 56 August 2013."

———. 2020. "National Health and Nutrition Examination Survey 2020." https://www.cdc.gov/nchs.

Shan, Zhilei, Colin D. Rehm, Gail Rogers, Mengyuan Ruan, Dong D. Wang, Frank B. Hu, Dariush Mozaffarian, Fang Fang Zhang, and Shilpa N. Bhupathiraju. 2019. "Trends in Dietary Carbohydrate, Protein, and Fat Intake and Diet Quality Among US Adults, 1999-2016." *JAMA - Journal of the American Medical Association* 322 (September): 1178–87. https://doi.org/10.1001/jama.2019.13771.

"The Third National Health and Nutrition Examination Survey (NHANES III, 1988-94) Reference Manuals and Reports-Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94-Analytic and Reporting Guidelines-Weighting and Estimation Methodology-Accounting for Item Nonresponse Bias-Field Operations Reference Manuals (by Topic) Next Page." 1996. http://www.cdc.gov/nchswww/nchshome.htm.

Appendix

```
knitr::opts_chunk$set(echo = FALSE , warning = FALSE, message = FALSE, comment = "")
library(aplore3)
library(tidyverse)
library(ggmosaic)
library(ggrepel)
```

```r
library(knitr)
library(summarytools)
library(patchwork)
library(binom)
library(DescTools)

DF <- nhanes

DF_1 <- DF[!(is.na(DF$wt)), ]

DF_1$strata <- as.factor(DF_1$strata)

DF_1$Age_Category <- cut(DF_1$age, breaks = c(0, 18, 35, 50, 65, Inf), labels = c('0 - 18','19 - 35','3

DF_1$Obesity <- ifelse(DF_1$obese == "Yes", 1, 0)

dfSummary(DF_1[,c(2:4, 6,7 )], style = "grid", plain.ascii = FALSE,
          varnumbers = FALSE, valid.col = FALSE, na.col = FALSE)

P1 <- ggplot(DF_1, aes(DF_1$tchol)) + geom_histogram() + theme_minimal() +
  labs( title = "Distribution of Total cholesterol (mg/dL)", x = "Total cholesterol (mg/dL)", y = "Coun

P2 <- ggplot(DF_1, aes(DF_1$hdl)) + geom_histogram() + theme_minimal() +
   labs(title = "Distribution of HDL- cholesterol (mg/dL)", x = "HDL- cholesterol (mg/dL)", y = "Count"

P3 <- ggplot(DF_1, aes(DF_1$sysbp)) + geom_histogram() + theme_minimal() +
  labs(title = "Distribution of Systolic blood pressure (mmHg)" , x = "Systolic blood pressure (mmHg)",

P4 <- ggplot(DF_1, aes(DF_1$dbp)) + geom_histogram() + theme_minimal() +
  labs( title = "Distribution of Diastolic blood pressure (mmHg)", x = "Diastolic blood pressure (mmHg)

(P1|P2 )/(P3|P4) + plot_annotation(title = "Distribution of Health Indicators")


P5 <- ggplot(DF_1, aes(DF_1$wt)) + geom_histogram() + theme_minimal() +
  labs( title = "Distribution of Weight (kg)", x = "Weight (kg)", y = "Count") + theme(plot.title = ele

P6 <- ggplot(DF_1, aes(DF_1$ht)) + geom_histogram() + theme_minimal() +
   labs(title = "Distribution of Standing Height (cm)", x = "Standing Height (cm)", y = "Count") + theme

P7 <- ggplot(DF_1, aes(DF_1$bmi)) + geom_histogram() + theme_minimal() +
  labs(title = "Distribution of Body mass index (kg/m2)" , x = "Body mass index (kg/m2)", y = "Count") +


(P5|P6)/(P7) + plot_annotation(title = "Distribution of BMI Variables")



P8 <- ggplot(na.omit(DF_1), aes(na.omit(DF_1)$vigwrk)) + geom_bar() + theme_minimal() +
  labs( title = "Composition of Vigorous work activity", x = "Response", y = "Count") + theme(plot.titl

P9 <- ggplot(na.omit(DF_1), aes(na.omit(DF_1)$modwrk)) + geom_bar() + theme_minimal() +
  labs( title = "Composition of Moderate work activity", x = "Response", y = "Count") + theme(plot.titl
```

```r
P10 <- ggplot(na.omit(DF_1), aes(na.omit(DF_1)$wlkbik)) + geom_bar() + theme_minimal() +
  labs( title = "Composition of Walk or Bicycle", x = "Response", y = "Count") + theme(plot.title = ele

P11 <- ggplot(na.omit(DF_1), aes(na.omit(DF_1)$vigrecexr)) + geom_bar() + theme_minimal() +
  labs( title = "Composition of Vigorous Recreational activities", x = "Response", y = "Count") + theme

P12 <- ggplot(na.omit(DF_1), aes(na.omit(DF_1)$modrecexr)) + geom_bar() + theme_minimal() +
  labs( title = "Composition of Moderate Recreational activities", x = "Response", y = "Count") + theme


P13 <- ggplot(na.omit(DF_1), aes(na.omit(DF_1)$sedmin)) + geom_histogram() + theme_minimal() +
  labs(title = "Distribution of  Minutes of Sedentary activity per Week" , x = " Minutes of Sedentary a


(P8|P9|P10)/(P11|P12) + plot_annotation(title = "Distribution of Activities")



P13


Obesity_Df <- as.data.frame(table(DF_1$obese)%>% prop.table())

ggplot(Obesity_Df, aes(x = Var1, y = Freq)) +
  geom_bar(stat = "identity", position = "dodge")  +
  labs(title = "Composition of People with Obesity and not ",x = "Obese or not", y = "Freq") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  geom_text(aes(label=round(Freq, 2)), vjust=-0.3, position = position_dodge(0.9), size=3.5)



ll.binom=function(pi,n,y){
  phat=y/n
  return(2*(y*log(phat/pi)+(n-y)*log((1-phat)/(1-pi))))
}

ll.binom.scaled=function(pi,n,y,alpha=0.05,df=1){
  quant=qchisq(alpha,df,lower=F)
  return(ll.binom(pi,n,y)-quant)
}

curve(ll.binom(x,n=length(DF_1$Obesity),y=sum(DF_1$Obesity)),from=0.14,to=0.17, xlab="p",
      ylab=paste0("l(p)"))
abline(h=3.84,lty=2,col="red")


lower=uniroot(ll.binom.scaled,interval=c(0.1,0.15),
              n=length(DF_1$Obesity),y=sum(DF_1$Obesity))
upper=uniroot(ll.binom.scaled,interval=c(0.16,0.165),
              n=length(DF_1$Obesity),y=sum(DF_1$Obesity))
abline(v=c(lower$root,upper$root), lty=2,col="blue")
text(0.15,12,labels=paste0("CI=[",
```

```
                        round(lower$root,2),",",
                        round(upper$root,2),"]"))

binom.confint(sum(DF_1$Obesity), length(DF_1$Obesity), conf.level = 0.95)
attach(DF_1)

Rev(table(gender, obese)) %>% prop.table(margin = 1)
Rev(table(gender, obese))%>%
  chisq.test()

Rev(table(gender, obese))
Rev(table(gender, obese)) %>% RelRisk( conf.level = 0.95,
                                                        method = "wald")

Rev(table( Age_Category, obese) )%>%
  prop.table(margin = 1)
Rev(table( Age_Category, obese) )%>%
  chisq.test()

Age_Gender_Obese <- DF_1 %>%
  group_by(Age_Category, gender) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count)


Age_Gender_Obese$Low_CI <- qbeta(0.05/2,shape1=Age_Gender_Obese$obesity_count,
                                 shape2=Age_Gender_Obese$total_count - Age_Gender_Obese$obesity_count+1)

Age_Gender_Obese$High_CI <- qbeta(1-0.05/2,shape1=Age_Gender_Obese$obesity_count+1,
                                  shape2=Age_Gender_Obese$total_count - Age_Gender_Obese$obesity_count)


# Plot the data
ggplot(Age_Gender_Obese, aes(x = Age_Category, y = proportion, fill = gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(x = "Age Category", y = "Proportion of Obesity", fill = "Gender") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + geom_text_repel(aes(label=round(proportion



Rev(table(vigwrk, obese))
Rev(table(vigwrk, obese)) %>% RelRisk( conf.level = 0.95,
                                                        method = "wald")

Rev(table(modwrk, obese))
Rev(table(modwrk, obese)) %>% RelRisk( conf.level = 0.95, method = "wald")

Rev(table(wlkbik, obese))

Rev(table(wlkbik, obese)) %>% RelRisk( conf.level = 0.95, method = "wald")
```

```r
Rev(table(vigrecexr, obese))

Rev(table(vigrecexr, obese)) %>% RelRisk( conf.level = 0.95, method = "wald")

Rev(table(modrecexr, obese))

Rev(table(modrecexr, obese)) %>% RelRisk( conf.level = 0.95, method = "wald")

detach(DF_1)
## Vigorous Work

vigwrk_Obese <- DF_1 %>%
  group_by(vigwrk) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


vigwrk_Obese$Low_CI <- qbeta(0.05/2,shape1=vigwrk_Obese$obesity_count,
                                shape2=vigwrk_Obese$total_count - vigwrk_Obese$obesity_count+1)

vigwrk_Obese$High_CI <- qbeta(1-0.05/2,shape1=vigwrk_Obese$obesity_count+1,
                                 shape2=vigwrk_Obese$total_count - vigwrk_Obese$obesity_count)



W1 <- ggplot(vigwrk_Obese, aes(x = vigwrk, y = proportion, fill = vigwrk)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = "Vigorous Work", x = "Vigorous Work", y = "Proportion of Obesity", fill = "Vigorous Work
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") + geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position


## Moderate Work

modwrk_Obese <- DF_1 %>%
  group_by(modwrk) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


modwrk_Obese$Low_CI <- qbeta(0.05/2,shape1=modwrk_Obese$obesity_count,
                                shape2=modwrk_Obese$total_count - modwrk_Obese$obesity_count+1)

modwrk_Obese$High_CI <- qbeta(1-0.05/2,shape1=modwrk_Obese$obesity_count+1,
                                 shape2=modwrk_Obese$total_count - modwrk_Obese$obesity_count)



W2 <- ggplot(modwrk_Obese, aes(x = modwrk, y = proportion, fill = modwrk)) +
```

```
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = " Moderate Work", x = "Moderate Work", y = "Proportion of Obesity", fill = "Moderate Work
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1 ,size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") + geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position
```

```
## Walking or Biking

wlkbik_Obese <- DF_1 %>%
  group_by(wlkbik) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


wlkbik_Obese$Low_CI <- qbeta(0.05/2,shape1=wlkbik_Obese$obesity_count,
                                shape2=wlkbik_Obese$total_count - wlkbik_Obese$obesity_count+1)

wlkbik_Obese$High_CI <- qbeta(1-0.05/2,shape1=wlkbik_Obese$obesity_count+1,
                                 shape2=wlkbik_Obese$total_count - wlkbik_Obese$obesity_count)


W3 <- ggplot(wlkbik_Obese, aes(x = wlkbik, y = proportion, fill = wlkbik)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = " Walking or Biking", x = "Walking or Biking", y = "Proportion of Obesity", fill = "Walki
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1 ,size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") +
  geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position = position_dodge(0.9), size=3.5)


## Vigorous Excerice

vigrecexr_Obese <- DF_1 %>%
  group_by(vigrecexr) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


vigrecexr_Obese$Low_CI <- qbeta(0.05/2,shape1=vigrecexr_Obese$obesity_count,
                                shape2=vigrecexr_Obese$total_count - vigrecexr_Obese$obesity_count+1)

vigrecexr_Obese$High_CI <- qbeta(1-0.05/2,shape1=vigrecexr_Obese$obesity_count+1,
                                 shape2=vigrecexr_Obese$total_count - vigrecexr_Obese$obesity_count)
```

```r
W4 <- ggplot(vigrecexr_Obese, aes(x = vigrecexr, y = proportion, fill = vigrecexr)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = " Vigorous Excercise",x = "Walking or Biking", y = "Proportion of Obesity", fill = "Vigo
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1 , size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") +
  geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position = position_dodge(0.9), size=3.5)



## Moderate Excerice

modrecexr_Obese <- DF_1 %>%
  group_by(modrecexr) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


modrecexr_Obese$Low_CI <- qbeta(0.05/2,shape1=modrecexr_Obese$obesity_count,
                                shape2=modrecexr_Obese$total_count - modrecexr_Obese$obesity_count+1)

modrecexr_Obese$High_CI <- qbeta(1-0.05/2,shape1=modrecexr_Obese$obesity_count+1,
                                 shape2=modrecexr_Obese$total_count - modrecexr_Obese$obesity_count)


W5 <- ggplot(modrecexr_Obese, aes(x = modrecexr, y = proportion, fill = modrecexr)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = " Moderate Excercise",x = "Moderate Excercise", y = "Proportion of Obesity", fill = "Mode
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1,  size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") +
  geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position = position_dodge(0.9), size=3.5)



(W1+W2+W3)/(W4+W5) + plot_annotation(title = "Proportion of Obesity in different Activity levels")


DF_1$Sedmin_Category <- cut(DF_1$sedmin, breaks = c(0, 10, 30, 60, 120, 240,360, Inf), labels = c('0 -

attach(DF_1)

Rev(table(Sedmin_Category, obese))%>% prop.table()

Rev(table(Sedmin_Category, obese))%>%
  chisq.test()
```

```r
detach(DF_1)

Sedmin_Category <- DF_1 %>%
  group_by(Sedmin_Category) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


Sedmin_Category$Low_CI <- qbeta(0.05/2,shape1=Sedmin_Category$obesity_count,
                                shape2=Sedmin_Category$total_count - Sedmin_Category$obesity_count+1)

Sedmin_Category$High_CI <- qbeta(1-0.05/2,shape1=Sedmin_Category$obesity_count+1,
                                 shape2=Sedmin_Category$total_count - Sedmin_Category$obesity_count)


ggplot(Sedmin_Category, aes(x = Sedmin_Category, y = proportion, fill = Sedmin_Category)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = "  Minutes of Sedentary activity per Week",x = " minutes of Sedentary activity per Week"
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1,  size = 8),
        axis.title.x = element_text(size = 10),
        axis.title.y = element_text(size = 10),
        legend.position="none") +
  geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position = position_dodge(0.9), size=3.5)




# Categorize Total Cholesterol levels
DF_1$Tchol_Category <- cut(DF_1$tchol,
                       breaks = c(-Inf, 200, 240, Inf),
                       labels = c("Normal", "Borderline High", "High"))

# Categorize HDL Cholesterol levels
DF_1$HDL_Category <- cut(DF_1$hdl,
                     breaks = c(-Inf, 40, 60, Inf),
                     labels = c("Low", "Normal", "High"))

# Categorize Systolic Blood Pressure levels
DF_1$Sysbp_Category <- cut(DF_1$sysbp,
                       breaks = c(-Inf, 120, 130, 140, Inf),
                       labels = c("Normal", "Elevated", "Hypertension Stage 1", "Hypertension Stage 2

# Categorize Diastolic Blood Pressure levels
DF_1$Dbp_Category <- cut(DF_1$dbp,
                       breaks = c(-Inf, 80, 90, Inf),
                       labels = c("Normal", "Elevated", "High"))


M1 <- ggplot(na.omit(DF_1)) + geom_mosaic(aes(x = product(Age_Category), fill = Tchol_Category)) + face
  theme_minimal() +
```

```
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(angle = 45, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") +  theme(plot.title = element_text(size = 9))

M2 <- ggplot(na.omit(DF_1)) + geom_mosaic(aes(x = product(Age_Category), fill = HDL_Category)) + facet_
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(angle = 45, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") +  theme(plot.title = element_text(size = 9))


M3 <- ggplot(na.omit(DF_1)) + geom_mosaic(aes(x = product(Age_Category), fill = Sysbp_Category)) + face
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(angle = 45, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") +  theme(plot.title = element_text(size = 9))

M4 <- ggplot(na.omit(DF_1)) + geom_mosaic(aes(x = product(Age_Category), fill = Dbp_Category)) + facet_
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(angle = 45, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") +  theme(plot.title = element_text(size = 9))


(M1+M2) + plot_annotation(title = "Distribution of Cholesterol Levels with Age and Gender")

(M3+M4) + plot_annotation(title = "Distribution of Blood Pressure Levels with Age and Gender")


# Perform Chi-Squared tests
tchol_test <- chisq.test(DF_1$Tchol_Category , DF_1$obese)
hdl_test <- chisq.test(DF_1$HDL_Category , DF_1$obese)
sysbp_test <- chisq.test(DF_1$Sysbp_Category , DF_1$obese)
dbp_test <- chisq.test(DF_1$Dbp_Category , DF_1$obese)

# Create a data frame to store the results
results <- data.frame(
  Variable = c("Total Cholesterol", "HDL Cholesterol", "Systolic Blood Pressure", "Diastolic Blood Pres
  Chi_Squared = c(tchol_test$statistic, hdl_test$statistic, sysbp_test$statistic, dbp_test$statistic),
  Degrees_of_Freedom = c(tchol_test$parameter, hdl_test$parameter, sysbp_test$parameter, dbp_test$parame
  P_Value = c(tchol_test$p.value, hdl_test$p.value, sysbp_test$p.value, dbp_test$p.value)
)

# Print the results
print(results)
```

```r
## TCL

Tchol_Category_Obese <- DF_1 %>%
  group_by(Tchol_Category) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


Tchol_Category_Obese$Low_CI <- qbeta(0.05/2,shape1=Tchol_Category_Obese$obesity_count,
                                     shape2=Tchol_Category_Obese$total_count - Tchol_Category_Obese$obesity_

Tchol_Category_Obese$High_CI <- qbeta(1-0.05/2,shape1=Tchol_Category_Obese$obesity_count+1 , shape2=Tch



H1 <- ggplot(Tchol_Category_Obese, aes(x = Tchol_Category, y = proportion, fill = Tchol_Category)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = "Total Cholesterol Level", x = "Total Cholesterol Level", y = "Proportion of Obesity") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") + geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position


## HDL

HDL_Category_Obese <- DF_1 %>%
  group_by(HDL_Category) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


HDL_Category_Obese$Low_CI <- qbeta(0.05/2,shape1=HDL_Category_Obese$obesity_count,
                                     shape2=HDL_Category_Obese$total_count - HDL_Category_Obese$obesity_cou

HDL_Category_Obese$High_CI <- qbeta(1-0.05/2,shape1=HDL_Category_Obese$obesity_count+1 , shape2=HDL_Cat



H2 <- ggplot(HDL_Category_Obese, aes(x = HDL_Category, y = proportion, fill = HDL_Category)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = "HDL Cholesterol Level", x = "HDL Cholesterol Level", y = "Proportion of Obesity") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") + geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position


## Sysbp
```

```r
Sysbp_Category_Obese <- DF_1 %>%
  group_by(Sysbp_Category) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


Sysbp_Category_Obese$Low_CI <- qbeta(0.05/2,shape1=Sysbp_Category_Obese$obesity_count,
                                     shape2=Sysbp_Category_Obese$total_count - Sysbp_Category_Obese$obesity_

Sysbp_Category_Obese$High_CI <- qbeta(1-0.05/2,shape1=Sysbp_Category_Obese$obesity_count+1 , shape2=Sysb



H3 <- ggplot(Sysbp_Category_Obese, aes(x = str_wrap( Sysbp_Category, width = 5), y = proportion, fill =
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = " Systolic blood pressure", x = "Systolic blood pressure Level", y = "Proportion of Obesi
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") + geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position


## DBP

Dbp_Category_Obese <- DF_1 %>%
  group_by(Dbp_Category) %>%
  summarise(obesity_count = sum(Obesity), total_count = n(),
            proportion = obesity_count / total_count) %>% na.omit()


Dbp_Category_Obese$Low_CI <- qbeta(0.05/2,shape1=Dbp_Category_Obese$obesity_count,
                                     shape2=Dbp_Category_Obese$total_count - Dbp_Category_Obese$obesity_cou

Dbp_Category_Obese$High_CI <- qbeta(1-0.05/2,shape1=Dbp_Category_Obese$obesity_count+1 , shape2=Dbp_Cate



H4 <- ggplot(Dbp_Category_Obese, aes(x = Dbp_Category, y = proportion, fill = Dbp_Category)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = Low_CI, ymax = High_CI), width = 0.2, position = position_dodge(0.9)) +
  labs(title = " Diastolic blood pressure
", x = "Diastolic blood pressure Level", y = "Proportion of Obesity") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 1, size = 8),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8),
        legend.position="none") + geom_text_repel(aes(label=round(proportion, 2)), vjust=-0.3, position



(H1+H2)/(H3+H4) + plot_annotation(title = "Proportion of Obesity with respect to the Levels of Health I
```