From Data to Diagnosis: Predicting Diabetes

Eirene Michella Tjhan

2024-06-20

Eirene Michella Tjhan
2702256630
Group 8

1. INTRODUCTION

Dataset : Diabetes Patients Data
Source : https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset?resource=download

This report discusses the problem of using different medical measurements to predict diabetes in Pima Indian women. The patients in the dataset are female Pima Indians, who are at least 21 years old and are known to have one of the highest rates of Type 2 diabetes globally.

Structure
1. Intoduction
2. Data Description
+ a) Problem Identification
+ b) Description
3. Data Preprocessing
4. Data Exploration
+ a) Examine Descirptive Statistic
+ b) Frequency Distribution
+ c) Data Visualization
+ d) Data Anomalies
5. Statistical Analysis
6. Discussion
7. Conclusion
8. References

Context
Type 2 diabetes is a chronic condition characterized by high blood sugar levels due to the body's inability to use insulin effectively. More than 50% of people in the Pima Indian population have been diagnosed with this condition, making them significantly at risk. The increasing number of cases of diabetes has been correlated to environmental factors, changes in lifestyle, and hereditary factors. Understanding and predicting diabetes within this population can help in early diagnosis and effective management of the disease.

Importance and Relevance
1. Diabetes causes serious health issues, such as hypertension and kidney failure, in which, both of them are common among Pima Indians.
2. In order to manage and lower the risk of diabetes, early detection may help in the implementation of changes in lifestyle and prevention strategies.

Intended Audience
1. Researchers who are interested in understanding the factors that may cause to diabetes in Pima Indian women.
2. Anyone interested in the wider effects of diabetes and the health issues the Pima Indian population is facing.

Goals
1. What factors that cause diabetes?
2. What variables can be the indicators of diabetes?
3. Why should we be aware of diabetes and what insight can we gain?

Methods
1. Exploratory Data Analysis (EDA): Statistic summary (location, spread, shape), Data visualization (identify outliers and find relationship between variables)
2. Statistical Analytic: Shapiro-Wilk test (data normality) and Pearson Correlation (linear relationship)

Assumptions and Limitations
- Assumptions: Based on the variables in this dataset, there are two major factors that affect diabetes, lifestyle (BMI, SkinThickness) and heredity (DiabetesPedigreeFunction).
- Limitations: This analysis is limited to the variables provided in the dataset, while there are external factors that may also affect diabetes. Thus, the analytical conclusion might not apply to other populations.

Conclusion
Prediction based on the analysis we conducted is important because diabetes can lead to many damages in the body, such as hypertension, which has the potential to damage blood circulation throughout the body. When we predict and identify people at risk of diabetes, healthcare professionals can step in with early preventive method to reduce the chances of issues for example high blood pressure and heart problems that come with diabetes.This analysis can significantly improve the overall health outcomes and quality of life for individuals affected by diabetes.

2. DATA DESCRIPTION

a) PROBLEM IDENTIFICATION

The data used for this analysis is originally from the National Institute of Diabetes and Digestive and Kidney Diseases and is made available on Kaggle by Mehmet Akturk. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on some measurement that is included in the dataset as well. All patients in the dataset are females Pima Indian heritage with at least 21 years old.

Who is Pima Indians female? Pima Indians (of Arizona, USA) have one of the highest recorded prevalence rates (total number of cases of a disease in a given statistical population at a given time, divided by the number of individuals in that population) of Type 2 diabetes in the world. Studies have shown that more than 50% of the adult population has Type 2 diabetes.

What is Type 2 diabetes?
Type 2 diabetes is a chronic condition can be recognized by high blood sugar levels due to the body's inability to use insulin effectively, often paired with insufficient insulin production. Insulin is a hormone produced by pancreas that helps glucose enter cells to be used for energy. In type 2 diabetes, the body's cells become resistant to insulin's effects, and the pancreas cannot produce enough insulin to overcome this resistance. And this can lead the pancreas to produce less insulin.

There are several factors that took part of these problem, such as:
1. Genetic factors
- The studies show that having a first-degree relative with diabetes (Diabetes Pedigree Function), will significantly increase a person's risk to develop diabetes as well.
2. Environment and lifestyle changes

- Pima Indians depending their life food supply by traditional agricultural, but due to the environmental change (weather), it is more possible to get food form the retail stores, rather than planting food themselves.
- So we can say, they move on to more modern lifestyle food, mostly processed foods that have high rates of sugar, compared with the food they gain from agricultural.
- The modernization also affect the amount of physical activity of the Pima Indians. They tend to be more 'lazy' because they rely on retail food stores, rather than growing their food by hand.
- The physical activity may contribute as an important factor in weight control (prevent obesity) and therefore, the prevention of diabetes.

Side affect of diabetes on Pima Indias?
1. Hyperglycemia : The cause of diabetes is because the body either does not produce enough insulin (Type 1 diabetes) or cannot effectively use the insulin it produces (Type 2 diabetes), leading to elevated blood glucose levels.
2. Hypertension : People with diabetes are more likely to have high blood pressure (hypertension) due to insulin resistance (in which the studies have found similar issue in Pima Indians), inflammation, and vascular damage caused by high blood glucose levels.

In conclusion, historical data suggest that diabetes was rare among the Pima before the adoption of modern lifestyles and diets. This is bad because the Arizona Pimas have an extraordinarily high rate of kidney disease in a result of diabetes, while kidney failure is a leading cause of death.

Source : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4418458/

b) DESCRIPTION

Explanation of attributes (variables):
- **Pregnancies**: To express the Number of pregnancies
- **Glucose**: Plasma glucose concentration a 2 hours in an oral glucose tolerance test (OGTT) (mg/dL)
- **BloodPressure**: To express the Blood pressure measurement, Diastolic blood pressure (mm Hg)
- **SkinThickness**: To express the thickness of the skin, Triceps skin fold thickness (mm)
- **Insulin**: To express the Insulin level in blood, 2-Hour serum insulin (mu U/ml)
- **BMI**: To express the Body mass index (weight in kg/(height in m)^2)
- **DiabetesPedigreeFunction**: To express the Diabetes percentage
- **Age**: To express the age
- **Outcome**: To express the final result 1 is Yes and 0 is No

EXPLANATION
What is **Oral Glucose Tolerance Test** (OGTT)?
The OGTT is a diagnostic tool that measures how well the body processes glucose, in which can help to identify insulin resistance and diabetes. Insulin resistance is a condition where the body's cells become less responsive to insulin, often leading to elevated blood sugar levels. Insulin is crucial for regulating blood glucose levels by making it easier for cells to absorb glucose and use it as energy.

Distributions of 2-hours **blood glucose levels** in OGTT:
- Normal: Less than 140 mg/dL (7.8 mmol/L)
- Prediabetes: 140 to 199 mg/dL (7.8 to 11.0 mmol/L)
- Diabetes: 200 mg/dL (11.1 mmol/L) or higher

Distributions of diastolic **blood pressure**:
- Normal: Less than 80 mm Hg
- Elevated: Less than 80 mm Hg q - Hypertension Stage 1: 80-89 mm Hg
- Hypertension Stage 2: 90 mm Hg or higher
- Hypertensive Crisis: Higher than 120 mm Hg (requires immediate medical attention)

SkinThickness for female adults:
People with higher skin fold thickness indicates higher body fat which is a major risk factor of Type 2 diabetes due to obesity & insulin resistance that can lead to higher glucose level. - Low: < 10mm

- Average: 10-25mm
- High: > 25mm

Insulin:
Insulin concentration in the blood measured two hours after consuming the glucose drink in OGTT (16 to 166 U/mL).

BMI can indicates whether a female has obesity or not: - Underweight: less than 18.5 kg/m²
- Normal: between 18.5 and 24.9 kg/m²
- Overweight: between 25.0 and 29.9 kg/m²
- Obesity:30.0 kg/m² and above

DiabetesPedigreeFunction:
percentage of risk of developing diabetes based on family history.

3. DATA PREPROCESSING

```r
source("explorationFunction.R") # BasicSummary function
library(corrplot) # correlation plot
library(ggplot2) # bar plot, scatter plot
library(dplyr)
library(plotly)
tinytex::install_tinytex()
```

```r
# import CSV file
data <- read.csv("C:/Users/acer/Downloads/diabetes.csv")
df <- data.frame(data)
head(df)
```

```
##   Pregnancies Glucose BloodPressure SkinThickness Insulin  BMI
## 1           6     148            72            35       0 33.6
## 2           1      85            66            29       0 26.6
## 3           8     183            64             0       0 23.3
## 4           1      89            66            23      94 28.1
## 5           0     137            40            35     168 43.1
## 6           5     116            74             0       0 25.6
##   DiabetesPedigreeFunction Age Outcome
## 1                    0.627  50       1
## 2                    0.351  31       0
## 3                    0.672  32       1
## 4                    0.167  21       0
## 5                    2.288  33       1
## 6                    0.201  30       0
```

How many records do we have? How many variables?

a. str function

```r
str(df)
```

```
## 'data.frame':    768 obs. of  9 variables:
##  $ Pregnancies             : int  6 1 8 1 0 5 3 10 2 8 ...
##  $ Glucose                 : int  148 85 183 89 137 116 78 115 197 125 ...
##  $ BloodPressure           : int  72 66 64 66 40 74 50 0 70 96 ...
##  $ SkinThickness           : int  35 29 0 23 35 0 32 0 45 0 ...
##  $ Insulin                 : int  0 0 0 94 168 0 88 0 543 0 ...
##  $ BMI                     : num  33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
##  $ DiabetesPedigreeFunction: num  0.627 0.351 0.672 0.167 2.288 ...
##  $ Age                     : int  50 31 32 21 33 30 26 29 53 54 ...
##  $ Outcome                 : int  1 0 1 0 1 0 1 0 1 1 ...
```

EXPLANATION
1. the first line shows that 'data' is a data frame, with 768 observations (rows), and 9 variables (attributes/columns).
2. the second, third, fourth, fifth, sixth, ninth, and tenth line tells us that they are integer variable (whole numbers).
3. the seventh and eighth line tells us that they are num variables (both integer and floating-point number).
4. although the variable 'outcome' get identified as int variable, it classified as categorical variable because 1 = yes, 0 = no.

b. summary function

```r
summary(df)
```

```
##   Pregnancies        Glucose      BloodPressure    SkinThickness
##  Min.   : 0.000   Min.   :  0.0   Min.   :  0.00   Min.   : 0.00
##  1st Qu.: 1.000   1st Qu.: 99.0   1st Qu.: 62.00   1st Qu.: 0.00
##  Median : 3.000   Median :117.0   Median : 72.00   Median :23.00
##  Mean   : 3.845   Mean   :120.9   Mean   : 69.11   Mean   :20.54
##  3rd Qu.: 6.000   3rd Qu.:140.2   3rd Qu.: 80.00   3rd Qu.:32.00
##  Max.   :17.000   Max.   :199.0   Max.   :122.00   Max.   :99.00
##     Insulin           BMI        DiabetesPedigreeFunction      Age
##  Min.   :  0.0   Min.   : 0.00   Min.   :0.0780           Min.   :21.00
##  1st Qu.:  0.0   1st Qu.:27.30   1st Qu.:0.2437           1st Qu.:24.00
##  Median : 30.5   Median :32.00   Median :0.3725           Median :29.00
##  Mean   : 79.8   Mean   :31.99   Mean   :0.4719           Mean   :33.24
##  3rd Qu.:127.2   3rd Qu.:36.60   3rd Qu.:0.6262           3rd Qu.:41.00
##  Max.   :846.0   Max.   :67.10   Max.   :2.4200           Max.   :81.00
##     Outcome
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :0.000
##  Mean   :0.349
##  3rd Qu.:1.000
##  Max.   :1.000
```

EXPLANATION
For all of the numeric variables, it shows the Tukey's five-number result (sample minimum, lower quartile, sample median, upper quartile, sample maximum) and the mean value of each variable.

What type is each variable (numeric, categorical, logical)?

Categorical (logical):
- Outcome : Represents final result, where 1 indicates "Yes" and 0 indicates "No" for diabetes

Numerical:
- Pregnancies : Represents number of pregnancies
- Glucose : Represents glucose level in blood
- BloodPressure : Represents blood pressure measurement
- SkinThickness : Represents thickness of the skin
- Insulin : Represents insulin level in blood
- BMI : Represents body mass index
- DiabetesPedigreeFunction : Represents diabetes percentage (genetic probability of diabtes based on family history)
- Age : Represents age

How many unique values does each variable have?

```r
sapply(df, function(x) length(unique(x)))
```

```
##              Pregnancies                  Glucose            BloodPressure
##                       17                      136                       47
##            SkinThickness                  Insulin                      BMI
##                       51                      186                      248
## DiabetesPedigreeFunction                      Age                  Outcome
##                      517                       52                        2
```

EXPLANATION
Pregnancies: 17 unique values ranging from 0 to 17.
Glucose: 135 unique values ranging from 0 to 199.
BloodPressure: 47 unique values ranging from 0 to 122.
SkinThickness: 52 unique values ranging from 0 to 99.
Insulin: 186 unique values ranging from 0 to 846.
BMI: 248 unique values ranging from 0 to 67.1.
DiabetesPedigreeFunction: 517 unique values ranging from 0.078 to 2.42.
Age: 52 unique values ranging from 21 to 81.
Outcome: 2 unique values, either 0 or 1, indicating the absence or presence of diabetes, respectively.

Is there duplicated data?

```r
sum(duplicated(df))
```

```
## [1] 0
```

What value occurs most frequently, and how often does it occur?

```r
BasicSummary(df)
```

```
##                   variable    type levels topLevel topCount topFrac missFreq
## 1              Pregnancies integer     17        1      135   0.176        0
## 2                  Glucose integer    136       99       17   0.022        0
## 3            BloodPressure integer     47       70       57   0.074        0
## 4            SkinThickness integer     51        0      227   0.296        0
## 5                  Insulin integer    186        0      374   0.487        0
## 6                      BMI numeric    248       32       13   0.017        0
## 7 DiabetesPedigreeFunction numeric    517    0.254        6   0.008        0
## 8                      Age integer     52       22       72   0.094        0
## 9                  Outcome integer      2        0      500   0.651        0
##   missFrac
## 1        0
## 2        0
## 3        0
## 4        0
## 5        0
## 6        0
## 7        0
## 8        0
## 9        0
```

EXPLANATION
Or we can see from the BasicSummary from 'topLevel' and 'topCount' column.
- Pregnancies : 1
- Glucose : 99
- BloodPressure : 70 - SkinThickness : 0
- Insulin : 0
- BMI : 32
- DiabetesPedigreeFunction : 0.254
- Age : 22
- Outcome : 0

Are there missing observations? If so, how frequently does this occur?

```r
missing_value <- sapply(df, function(x) sum(is.na(x)))
missing_value
```

```
##              Pregnancies                  Glucose            BloodPressure
##                        0                        0                        0
##            SkinThickness                  Insulin                      BMI
##                        0                        0                        0
## DiabetesPedigreeFunction                      Age                  Outcome
##                        0                        0                        0
```

If we observe the dataset manually, we may found it weird there are so many 0 value. So in this dataset the missing value is not represent by N/A or NULL but with 0.

```r
sum(df$Pregnancies == 0)
```

```
## [1] 111
```

```r
sum(df$Glucose == 0)
```

```
## [1] 5
```

```r
sum(df$BloodPressure == 0)
```

```
## [1] 35
```

```r
sum(df$SkinThickness == 0)
```

```
## [1] 227
```

```r
sum(df$Insulin == 0)
```

```
## [1] 374
```

```r
sum(df$BMI == 0)
```

```
## [1] 11
```

```r
sum(df$DiabetesPedigreeFunction == 0)
```

```
## [1] 0
```

```r
sum(df$Age == 0)
```

```
## [1] 0
```

```r
sum(df$Outcome == 0)
```

```
## [1] 500
```

The possibility of variable Pregnancies, DiabetesPedigreeFunction, and Outcome to have 0 as their value is understandable, because in fact not all of them have pregnant. There's possibility they have 0 percentage of having diabetes. Outcome is categorical variable that act as the final result whether a person has diabetes or not (1 yes, 0 no).

```r
## Replace missing values with mean/median for numerical variable
median_SkinThickness <- median(df$SkinThickness)
median_Insulin <- median(df$Insulin)
mean_Glucose <- mean(df$Glucose)
mean_BloodPressure <- mean(df$BloodPressure)
mean_BMI <- mean(df$BMI)
```

```r
df$Glucose[df$Glucose==0] <- mean_Glucose
df$BloodPressure[df$BloodPressure==0] <- mean_BloodPressure
df$SkinThickness[df$SkinThickness==0] <- median_SkinThickness
df$Insulin[df$Insulin==0] <- median_Insulin
df$BMI[df$BMI==0] <- mean_BMI

sum(df$Glucose==0)
```

```
## [1] 0
```

```r
sum(df$BloodPressure==0)
```

```
## [1] 0
```

```r
sum(df$SkinThickness==0)
```

```
## [1] 0
```

```r
sum(df$Insulin==0)
```

```
## [1] 0
```

```r
sum(df$BMI==0)
```

```
## [1] 0
```

So the preprocessing of the dataset: replacing rows that have 0 in Glucose/BloodPressure/SkinThickness/Insulin/BMI with the median or mean of the variables itself.

4. DATA EXPLORATION

a) EXAMINE DESCRIPTIVE STATISTIC

Location (Central Tendency)

```r
# numerical (mean/median/variance)
# average value for each variable, representing the central tendency or location
mean_Pregnancies = mean(df$Pregnancies)
mean_Glucose = mean(df$Glucose)
mean_BloodPressure = mean(df$BloodPressure)
mean_SkinThickness = mean(df$SkinThickness)
mean_Insulin = mean(df$Insulin)
mean_BMI = mean(df$BMI)
mean_DiabetesPedigreeFunction = mean(df$DiabetesPedigreeFunction)
mean_Age = mean(df$Age)

mean_Pregnancies
```

```
## [1] 3.845052
```

```
mean_Glucose
```

```
## [1] 121.6816
```

```
mean_BloodPressure
```

```
## [1] 72.25481
```

```
mean_SkinThickness
```

```
## [1] 27.33464
```

```
mean_Insulin
```

```
## [1] 94.65234
```

```
mean_BMI
```

```
## [1] 32.45081
```

```
mean_DiabetesPedigreeFunction
```

```
## [1] 0.4718763
```

```
mean_Age
```

```
## [1] 33.24089
```

EXPLANATION
1. The mean frequencies of pregnancy occur is 3.85.
2. The average of Glucose (Plasma glucose concentration) level of (121.68 mg/dL) on Oral glucose tolerance tests (OGTT) shows that it was categorized as normal glucose level (Normal: Less than 140 mg/dL).
3. The average of diastolic blood pressure (72.25 mm Hg) shows that it was indeed normal (Normal: Less than 80 mm Hg).
4. The average of skinThickness (27.33mm) show that it was consider as High, in which can indicates higher body fat (obesity), but there might be other factor that can affect this (SkinThickness alone may not necessarily mean obesity).
5. The average of insulin concentration in the blood measured two hours after consuming the glucose drink in OGTT (94.65 U/mL) can be considered as low (normal 100-150) and may indicate the insulin resistance (problems of handling glucose).
6. The average BMI of (32.45 kg/m^2) shows that there are general obesity trend within the samples (obesity : 30.0 kg/m^2 and above), and this can lead to higher risk for diabetes. 7. The average of diabetesPedigreeFunction (0.47 or 47%) shows that there are significant heredity predisposition towards diabetes, in which may be a major factor in predicting the diabetes outcome considering heredity plays a huge part in diabetes.
8. The mean frequencies of age is around 33.

```r
# categorical
outcome_counts <- table(df$Outcome)
labels_outcome <- c("No Diabetes", "Has Diabetes")

outcome_pie <- plot_ly(labels = labels_outcome,
                       values = as.numeric(outcome_counts),
                       type = 'pie',
                       textinfo = 'label+percent') %>%
  layout(title = "Diabetes Distribution")

outcome_pie
```

There are 34.9% of adult female participants sample in Pima Indians of Arizona that has diabetes.

Spread

```r
# numerical(IQR/range)
# Function to calculate Q1, Q3, and IQR
calculate_quartiles <- function(x) {
  q1 <- quantile(x, 0.25, na.rm = TRUE)
  q3 <- quantile(x, 0.75, na.rm = TRUE)
  iqr <- IQR(x, na.rm = TRUE)
  return(c(Q1 = q1, Q3 = q3, IQR = iqr))
}

quartile_stats <- sapply(df, calculate_quartiles)
quartile_stats <- t(quartile_stats)
print(quartile_stats)
```

```
##                            Q1.25%    Q3.75%     IQR
## Pregnancies               1.00000   6.00000  5.0000
## Glucose                  99.75000 140.25000 40.5000
## BloodPressure            64.00000  80.00000 16.0000
## SkinThickness            23.00000  32.00000  9.0000
## Insulin                  30.50000 127.25000 96.7500
## BMI                      27.50000  36.60000  9.1000
## DiabetesPedigreeFunction  0.24375   0.62625  0.3825
## Age                      24.00000  41.00000 17.0000
## Outcome                   0.00000   1.00000  1.0000
```

EXPLANATION
1. With a Q1 of 1, a Q3 of 6, and an IQR of 5, the number of pregnancies shows that 50% of the data lies between 1 and 6 pregnancies.
2. The glucose levels have a Q1 of 99.75, a Q3 of 140.25, and an IQR of 40.5. Altho the data also shows that 50% of the participant's blood glucose levels in the OGTT remain considered as normal, the spread (IQR) indicates that there is medium variability in blood glucose levels among individuals.
3. The diastolic blood pressure measurements, which have a Q1 of 64, a Q3 of 80, and an IQR of 16, shows that although individual variations may exist in the data, 50% of the participant's blood pressure still considered as normal.
4. Skin thickness values show a narrow distribution in the data, with a range from a Q1 of 23 to a Q3 of 32 with an IQR of 9. However, 50% of participants have a skin thickness level that can be considered high,

showing higher levels of body fat.
5. Insulin levels have a Q1 of 30.5, a Q3 of 127.25, and an IQR of 96.75, showing there are significant variability that might be caused by outliers.
6. The body mass index has a Q1 of 27.5, a Q3 of 36.6, and an IQR of 9.1, indicating there are variability of body mass indices. However, 50% of participants are considered as Overweight and Obesed.
7. The Diabetes Pedigree Function has a Q1 of 0.24375, a Q3 of 0.62625, and an IQR of 0.3825, showing there are medium variability in the percentage of risk of developing diabetes.
8. The age distribution shows a Q1 of 24, a Q3 of 41, and an IQR of 17, showing that there is a wide age range among individuals.
9. The outcome variable, which is categorical of (0 = non-diabetical) and (1 = diabetical), has a Q1 of 0, a Q3 of 1, and an IQR of 1.

Shape

```r
# graphical (histogram)

create_histogram <- function(data, variable_name){
  plot <- plot_ly(
    data = data,
    x = ~get(variable_name),
    type = 'histogram',
    textinfo = "text") %>%
    layout(
      title = paste("Histogram of", variable_name),
      xaxis = list(title = variable_name),
      yaxis = list(title = "Frequency")
    )
  return(plot)
}

histograms <- lapply(names(df), function(var) create_histogram(df, var))
for(plot in histograms){
  print(plot)
}
```

EXPLANATION
From the histogram above we can conclude that:
- Pregnancies : right (positive) skewed
- Glucose : almost normally distributed but right (positive) skewed
- BloodPressure : normally distributed
- SkinThickness : right (positive) skewed
- Insulin : right (positive) skewed
- BMI : right (positive) skewed
- DiabetesPedigreeFunction : right (positive) skewed
- Age : right (positive) skewed
- Outcome : right (positive) skewed

```r
# Function to calculate skewness
calculate_skewness <- function(x){
  n <- length(x)
  mean_x <- mean(x, na.rm = TRUE)
  sd_x <- sd(x, na.rm = TRUE)
```

```r
  skewness <- sum((x - mean_x)^3, na.rm = TRUE) / (n * sd_x^3)
  return(skewness)
}

# Function to calculate kurtosis
calculate_kurtosis <- function(x){
  n <- length(x)
  mean_x <- mean(x, na.rm = TRUE)
  sd_x <- sd(x, na.rm = TRUE)
  kurtosis <- sum((x - mean_x)^4, na.rm = TRUE) / (n * sd_x^4) - 3
  return(kurtosis)
}

skewness_results <- sapply(df, function(x) if (is.numeric(x)) calculate_skewness(x) else NA)
kurtosis_results <- sapply(df, function(x) if (is.numeric(x)) calculate_kurtosis(x) else NA)

results <- data.frame(
  Skewness = skewness_results,
  Kurtosis = kurtosis_results
)

print(results)
```

```
##                           Skewness   Kurtosis
## Pregnancies              0.8981549  0.1421840
## Glucose                  0.5311437 -0.2720579
## BloodPressure            0.1723748  1.0538400
## SkinThickness            1.2182837  4.6585559
## Insulin                  2.6826696  9.6369965
## BMI                      0.5987572  0.8973587
## DiabetesPedigreeFunction 1.9124179  5.5285389
## Age                      1.1251880  0.6217269
## Outcome                  0.6325383 -1.6019762
```

EXPLANATION Pregnancies
- Skewness: 0.898 (positively skewed or it means right tailed distribution)
- Kurtosis: 0.142 (normal distribution)

Glucose
- Skewness: 0.531 (positively skewed or right tailed distribution)
- Kurtosis: -0.272 (slightly lighter tails or lesser outlier)

BloodPressure
- Skewness: 0.172 (symmetric)
- Kurtosis: 1.054 (slightly heavier tails or more outlier)

SkinThickness
- Skewness: 1.218 (positively skewed or right tailed distribution)
- Kurtosis: 4.659 (significantly heavy tails or more outlier)

Insulin
- Skewness: 2.683 (positively skewed or right tailed distribution)
- Kurtosis: 9.637 (very heavy tails or there so many outlier)

BMI
- Skewness: 0.599 (positively skewed or right tailed distribution)
- Kurtosis: 0.897 (slightly heavier tails or more outllier)

DiabetesPedigreeFunction

- Skewness: 1.912 (positively skewed or right tailed outlier)
- Kurtosis: 5.529 (significantly heavy tails or more outlier)

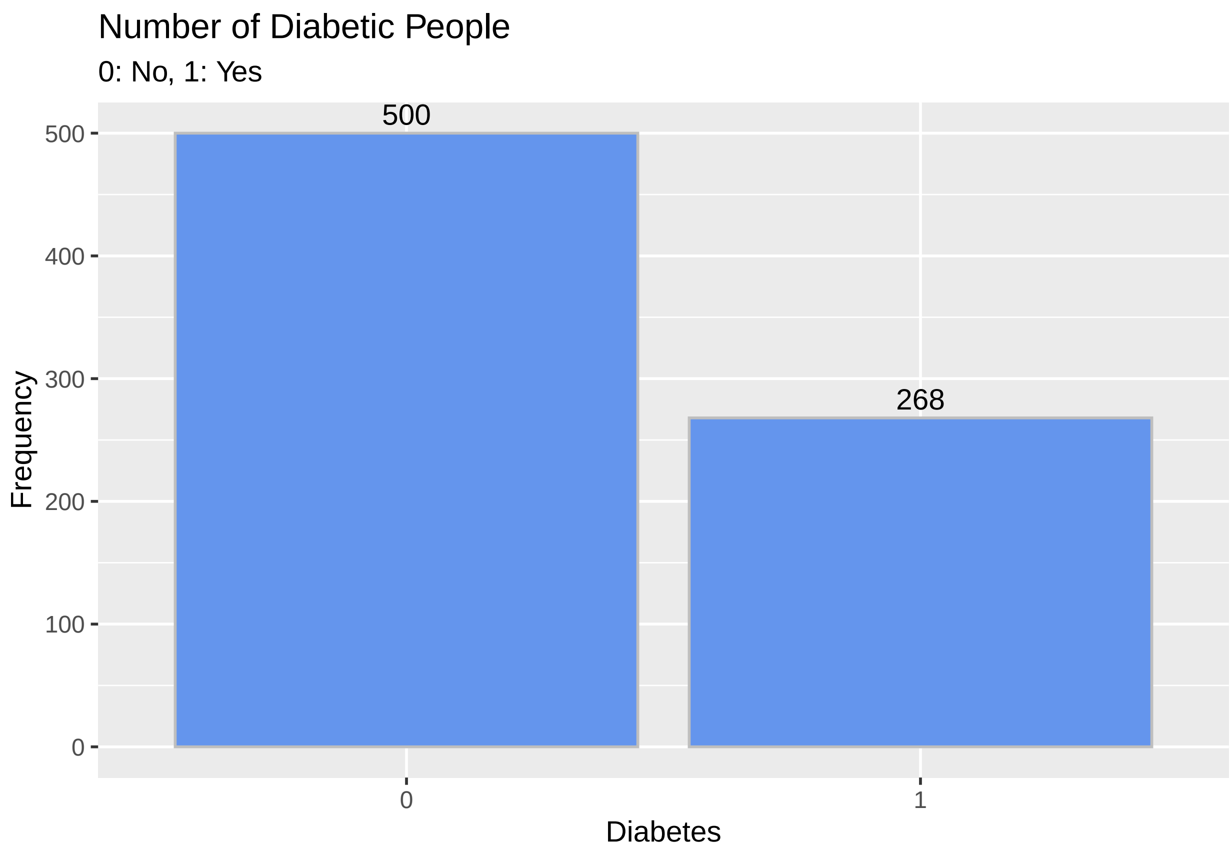
Age
- Skewness: 1.125 (positively skewed or right tailed distributiom)
- Kurtosis: 0.622 (slightly heavier tails or more outlier)

Outcome
- Skewness: 0.633 (spositively skewed or right tailed distribution)
- Kurtosis: -1.602 (significantly lighter tails or lesser outlier)

b) FREQUENCY DISTRIBUTION

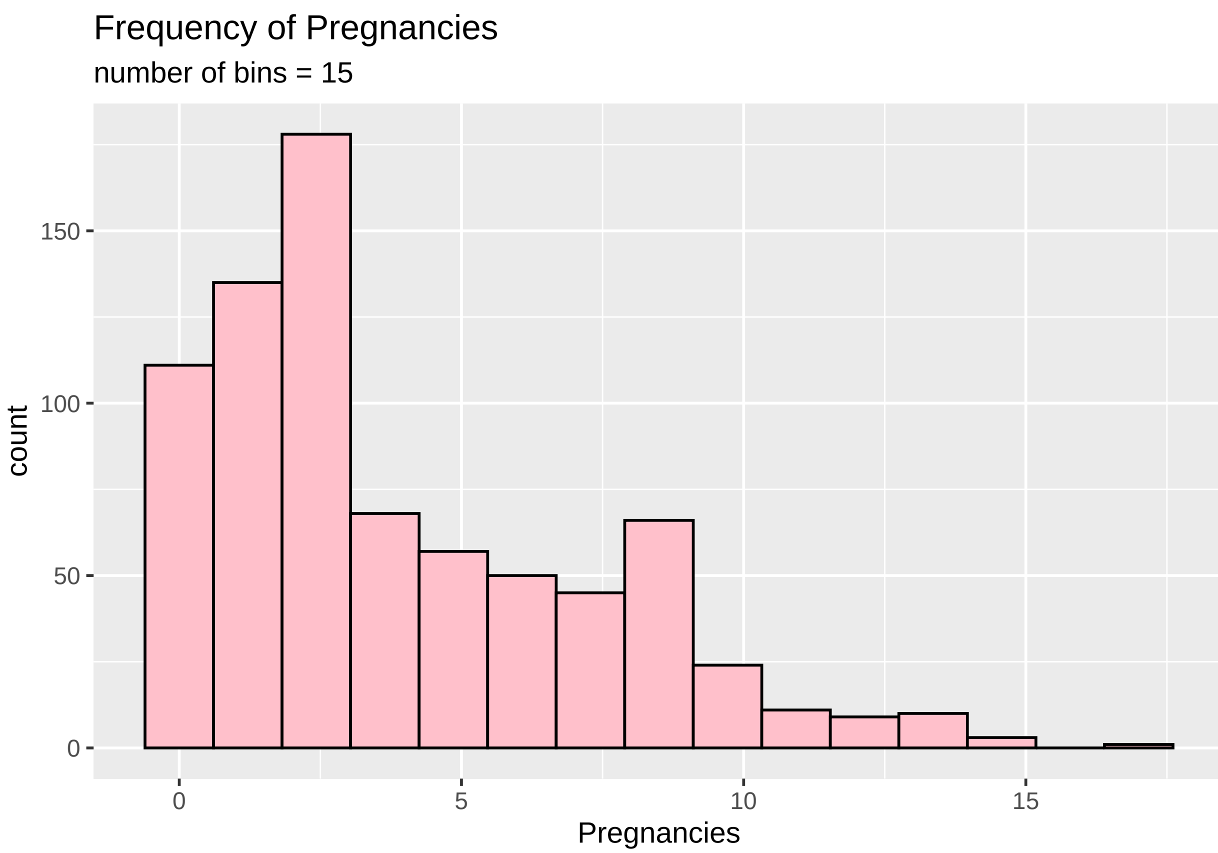
Number of Diabetic People

```
plotdata <- df %>%
 count(Outcome)

ggplot(plotdata,
       aes(x = reorder(Outcome, -n), y = n)) +
  geom_bar(stat="identity", fill = "cornflowerblue",color="gray") + geom_text(aes(label = n), vjust=-0.
  labs(x = "Diabetes",
       y = "Frequency",
       title  = "Number of Diabetic People",
       subtitle = "0: No, 1: Yes")
```



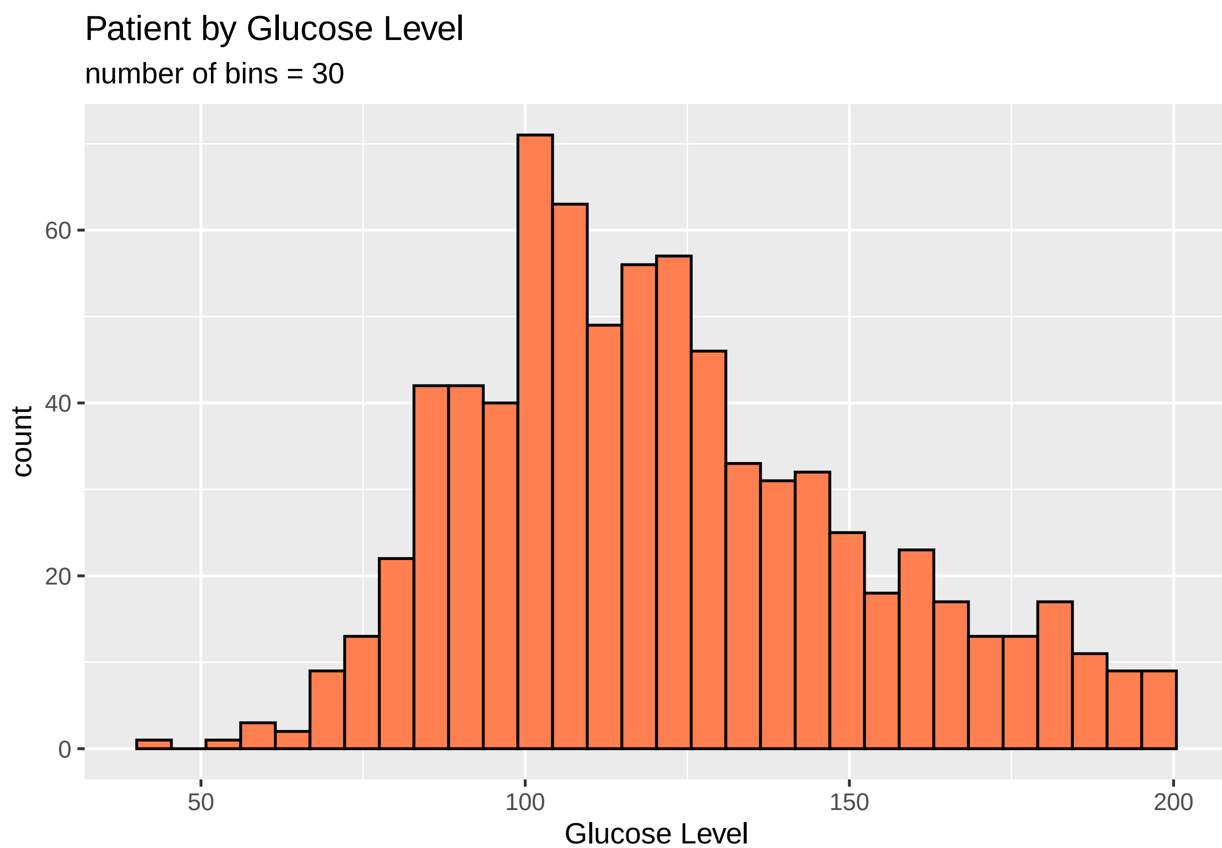
Frequency of Pregnancies

```r
ggplot(df, aes(x = Pregnancies)) +
  geom_histogram(fill = "pink",
                 color = "black",
                 bins = 15) +
  labs(title="Frequency of Pregnancies",
       subtitle = "number of bins = 15",
       x = "Pregnancies")
```



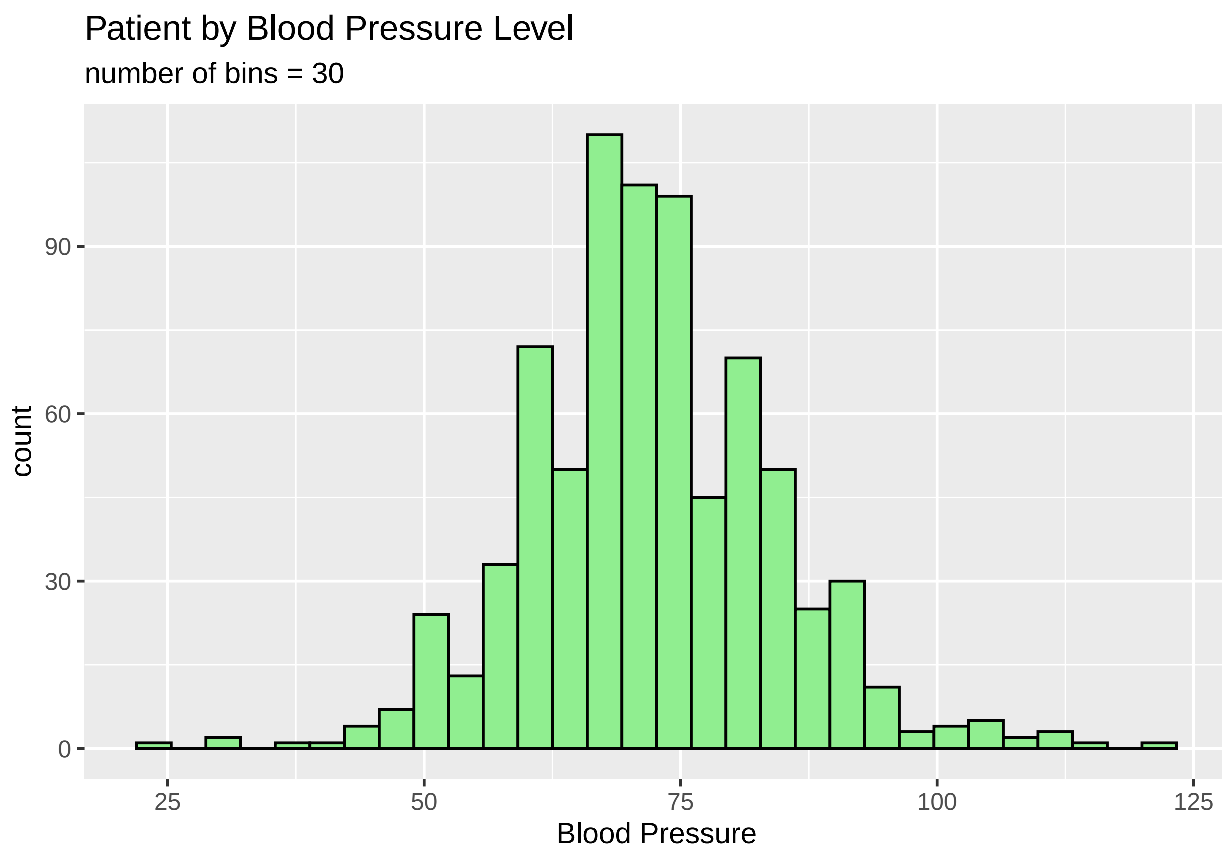
Glucose level

```r
ggplot(df, aes(x = Glucose)) +
  geom_histogram(fill = "coral",
                 color = "black",
                 bins = 30) +
  labs(title="Patient by Glucose Level",
       subtitle = "number of bins = 30",
       x = "Glucose Level")
```



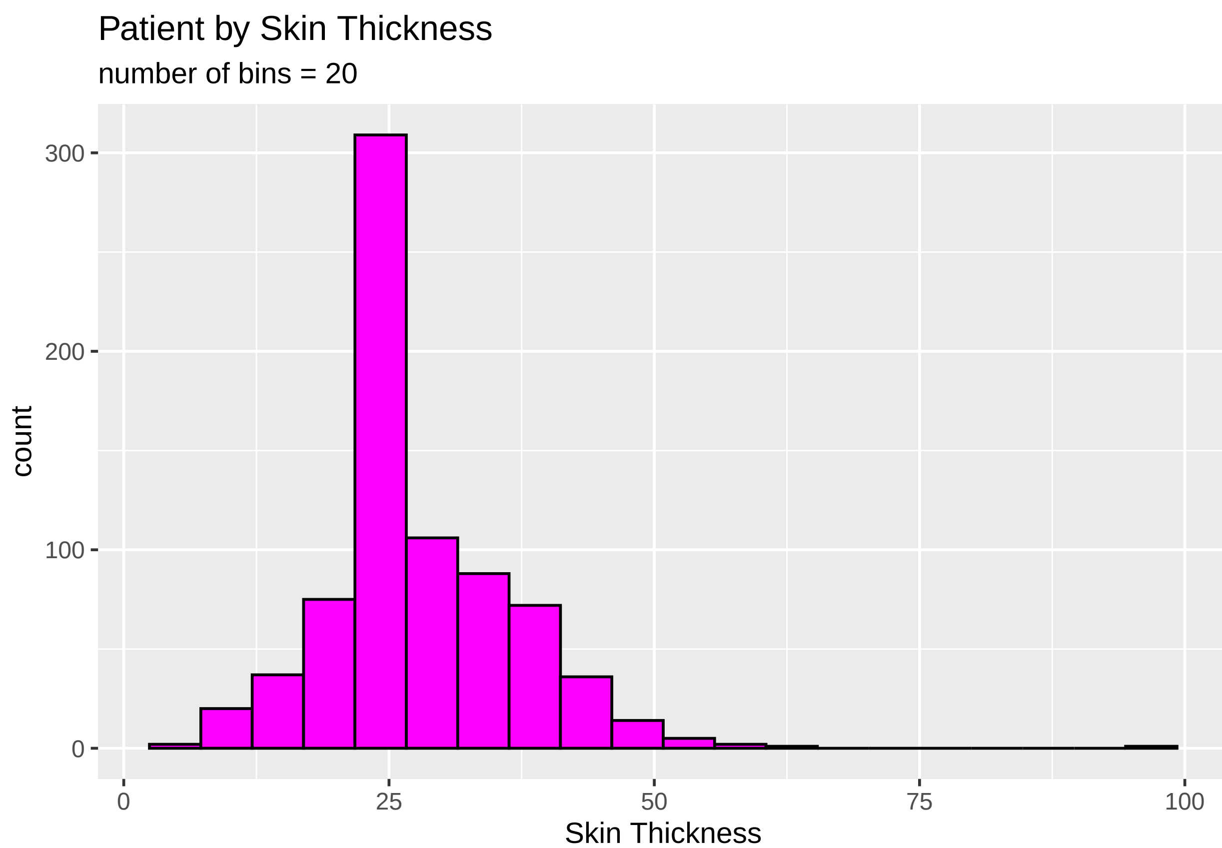
Blood Pressure level

```
ggplot(df, aes(x = BloodPressure)) +
  geom_histogram(fill = "lightgreen",
                 color = "black",
                 bins = 30) +
  labs(title="Patient by Blood Pressure Level",
       subtitle = "number of bins = 30",
       x = "Blood Pressure")
```



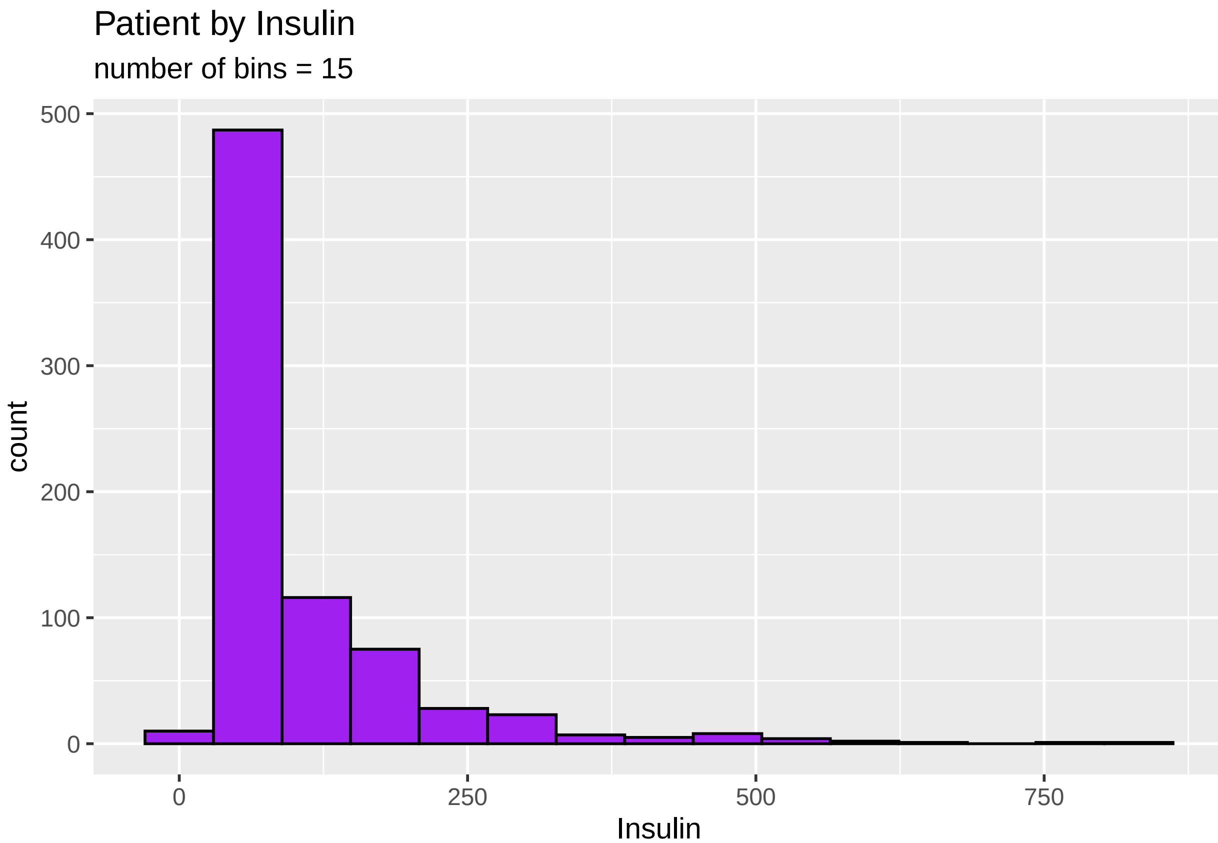
Participant by Skin Thickness

```
ggplot(df, aes(x = SkinThickness)) +
  geom_histogram(fill = "magenta",
                 color = "black",
                 bins = 20) +
  labs(title="Patient by Skin Thickness",
       subtitle = "number of bins = 20",
       x = "Skin Thickness")
```



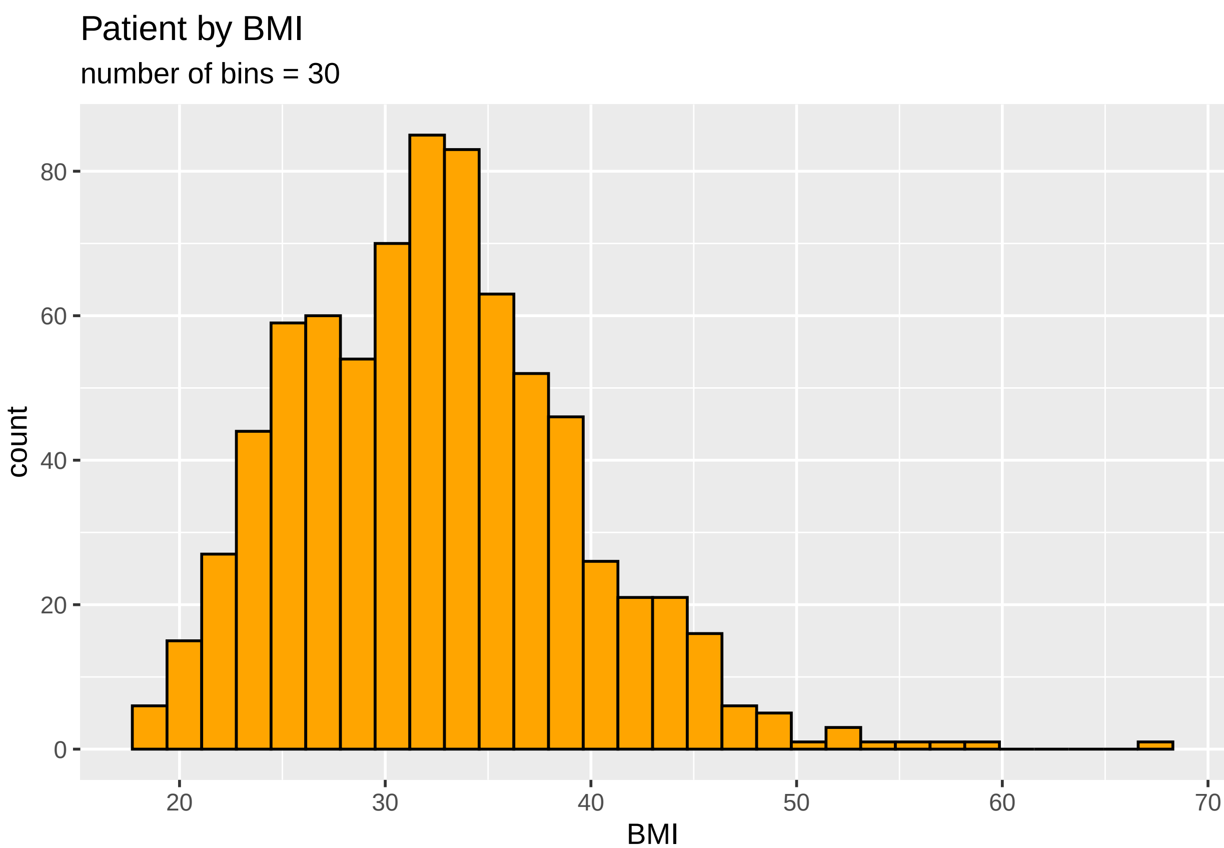
Participant by Insulin

```r
ggplot(df, aes(x = Insulin)) +
  geom_histogram(fill = "purple",
                 color = "black",
                 bins = 15) +
  labs(title="Patient by Insulin",
       subtitle = "number of bins = 15",
       x = "Insulin")
```



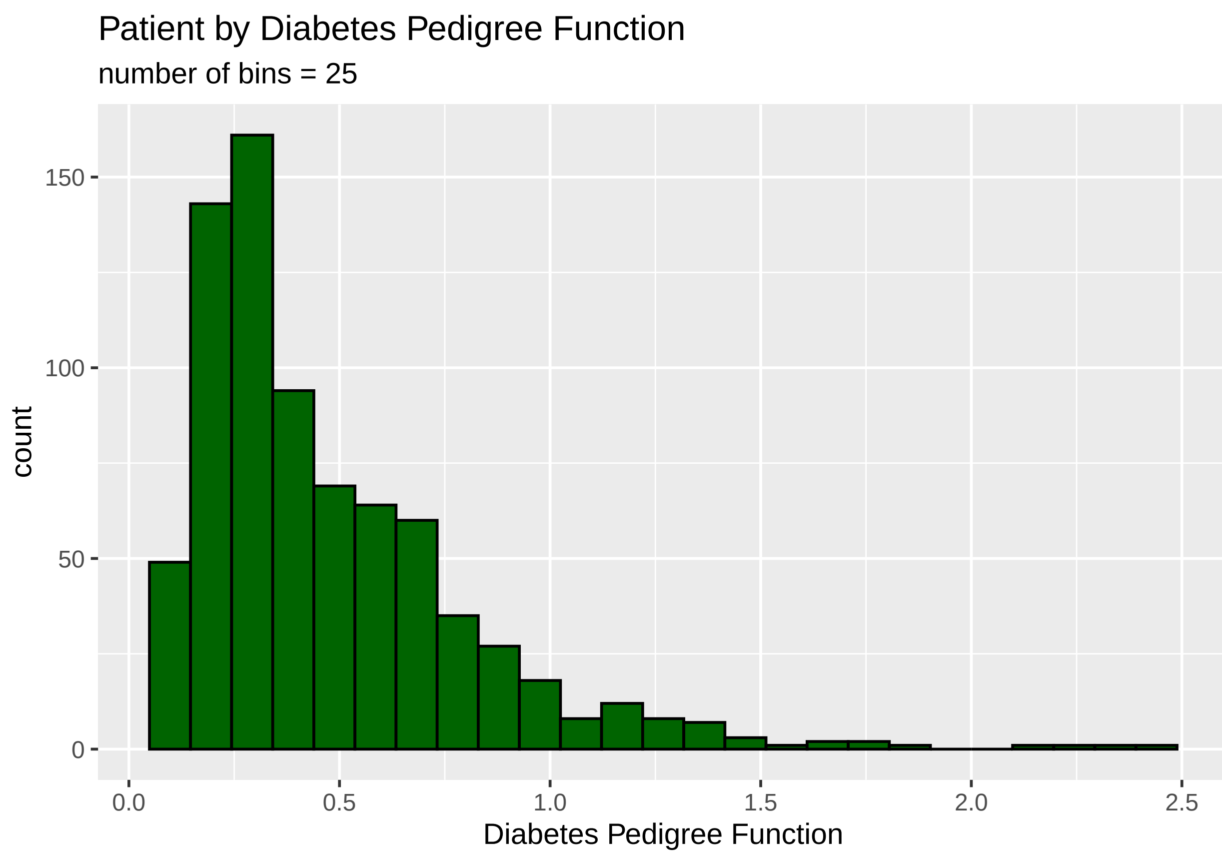
Participant by BMI

```
ggplot(df, aes(x = BMI)) +
  geom_histogram(fill = "orange",
                 color = "black",
                 bins = 30) +
  labs(title="Patient by BMI",
       subtitle = "number of bins = 30",
       x = "BMI")
```



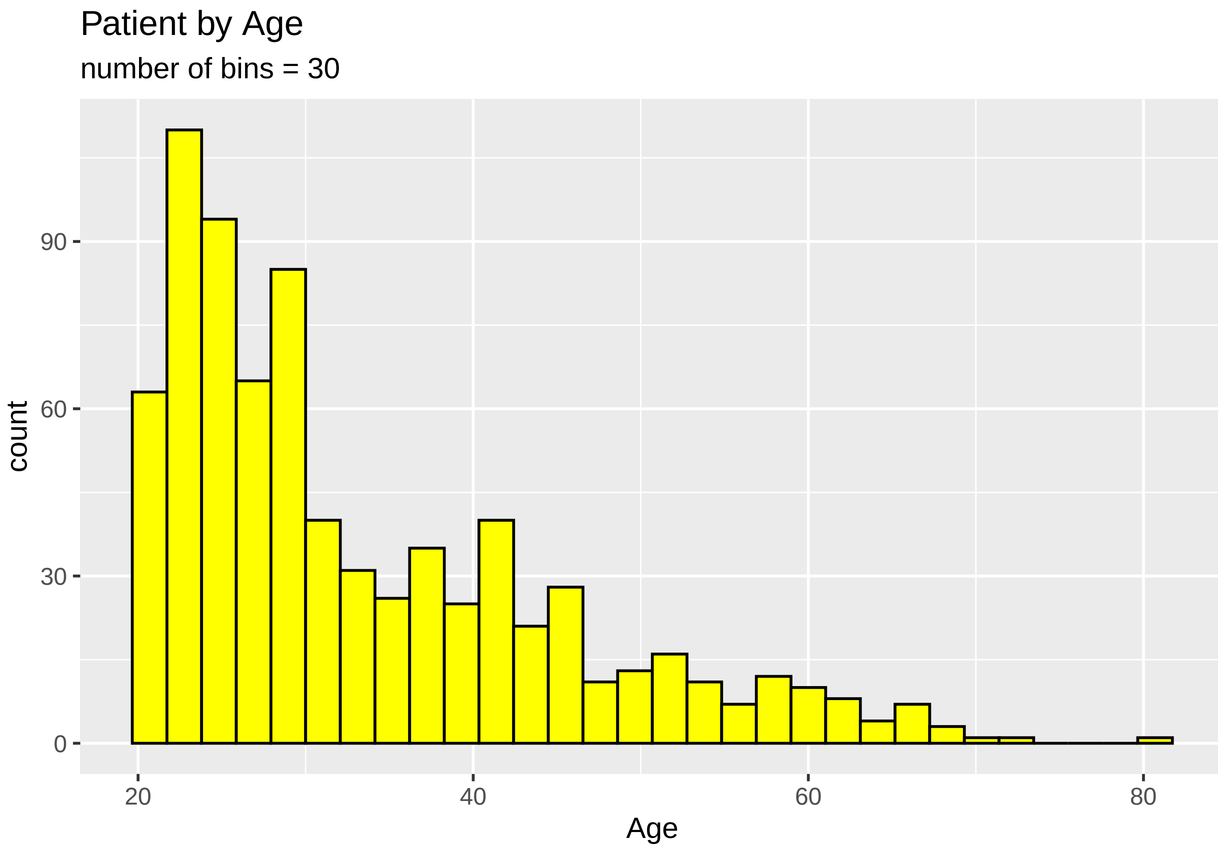
Diabetes Pedigree Function

```
ggplot(df, aes(x = DiabetesPedigreeFunction)) +
  geom_histogram(fill = "darkgreen",
                 color = "black",
                 bins = 25) +
  labs(title="Patient by Diabetes Pedigree Function",
       subtitle = "number of bins = 25",
       x = "Diabetes Pedigree Function")
```



Participant by Age

```
ggplot(df, aes(x = Age)) +
  geom_histogram(fill = "yellow",
                 color = "black",
                 bins = 30) +
  labs(title="Patient by Age",
       subtitle = "number of bins = 30",
       x = "Age")
```



c) DATA VISUALIZATION

What is the indicator of Diabetes?

Glucose vs Outcome

```
plot_ly(data = df, x = ~Outcome, y = ~Glucose, type = "box", boxmean = TRUE,
        marker = list(color = 'red'),
        line = list(color = 'red'),
        fillcolor = 'rgba(255,0,0,0.5)') %>%
  layout(title = "Glucose Levels vs Outcome",
         xaxis = list(title = "Outcome"),
         yaxis = list(title = "Glucose Level"))
```

As mentioned before, the dataset contains the result of Oral Glucose Tolerance Test (OGTT), in which is used to observe the body's reaction in processing glucose, which can help to identify insulin resistance and diabetes.

Here in the plot, we can see median glucose level is significantly higher for individuals with diabetes (Outcome = 1) compared to those without diabetes (Outcome = 0).
- Median glucose level without diabetes = 107.5
- Median glucose level with diabetes = 140
- Mean glucose level without diabetes = 110.7
- Mean glucose level with diabetes = 142.2

By looking the outliers between each group, we can see that the diabetic group glucose levels are closer to a normal distribution, while the non-diabetic group glucose level are right-skewed. The diabetical group has fewer outliers showing that they have more consistency in their range of glucose level.

The IQR is higher for individuals with diabetes, indicating there is more variability in glucose levels within the diabetic group (IQR = 119 ~ 167), while the non-diabetic group has more narrow variation (IQR = 93 ~ 125).

In conclusion, the box plot clearly shows that the result of OGTT for individuals with diabetes, tend to keep their glucose levels high even after 2 hours, showing there are insulin incapability of regulate glucose concentration in blood, this condition is also known as Insulin Resistance.

Insulin vs Outcome

```
plot_ly(data = df, x = ~Outcome, y = ~Insulin, type = "box", boxmean = TRUE,
        marker = list(color = 'red'),
        line = list(color = 'red'),
        fillcolor = 'rgba(255,0,0,0.5)') %>%
  layout(title = "Insulin Levels vs Outcome",
         xaxis = list(title = "Outcome"),
         yaxis = list(title = "Insulin Level"))
```

Here in the plot, we can see median glucose level is higher for individuals without diabetes (Outcome = 0) compared to those with diabetes (Outcome = 1).
- Median glucose level without diabetes = 39
- Median glucose level with diabetes = 30.5
- Mean glucose level without diabetes = 83.188
- Mean glucose level with diabetes = 116.041
By looking the outliers between each group, we can see that both the diabetic and non-diabetical group insulin levels have many extreme values, showing that there are individuals with or without diabetes that tend to have a high insulin level postest.

The IQR is higher for individuals with diabetes, indicating there is more variability in inslulin levels within the diabetic group (IQR = 30.5 ~ 167.5), while the non-diabetic group has more narrow variation (IQR = 30.5 ~ 105).

In conclusion, the box plot shows that the result of OGTT for people with diabetes, the insulin levels remains high or show a delayed response, indicating the failed body's attempt to manage the high blood glucose levels. That makes insulin can't really get the job done breaking down glucose.

How does lifestyle affect Diabetes?

SkinThickness vs BMI vs Outcome

```
plot_ly(
  data = df,
  x = ~SkinThickness,
  y = ~BMI,
  type = "scatter",
  mode = "markers",
  color = ~Outcome,
  colors = c('Non Diabetes' = 'yellow', 'Diabetes' = 'red')
) %>%
  layout(
    title = "Skin Thickness vs BMI",
    xaxis = list(title = "Skin Thickness"),
```

```
    yaxis = list(title = "BMI"),
    legend = list(title = list(text = 'Outcome'))
  )
```

On the plot we can see there a positive linear relationship between BMI and SkinThickness. Having a thicker skin fold indicates a higher body fat, people that have higher BMI will also have a high skinfold thickness.

In the plot we can also see that individuals that in fact have a higher BMI and higher levels of SkinThickness are more likely to suffers from diabetes, due to obesity.

```
df.temp <- df
df.temp$BMI[df$BMI<18.5] <- "Underweight"
df.temp$BMI[df$BMI>=18.5 & df$BMI<=24.9] <- "Normal"
df.temp$BMI[df$BMI>=25.0 & df$BMI<=29.9] <- "Overweight"
df.temp$BMI[df$BMI>=30.0] <- "Obesity"
df.temp
```

```
##     Pregnancies  Glucose BloodPressure SkinThickness Insulin         BMI
## 1             6 148.0000      72.00000            35    30.5     Obesity
## 2             1  85.0000      66.00000            29    30.5  Overweight
## 3             8 183.0000      64.00000            23    30.5      Normal
## 4             1  89.0000      66.00000            23    94.0  Overweight
## 5             0 137.0000      40.00000            35   168.0     Obesity
## 6             5 116.0000      74.00000            23    30.5  Overweight
## 7             3  78.0000      50.00000            32    88.0     Obesity
## 8            10 115.0000      69.10547            23    30.5     Obesity
## 9             2 197.0000      70.00000            45   543.0     Obesity
## 10            8 125.0000      96.00000            23    30.5     Obesity
## 11            4 110.0000      92.00000            23    30.5     Obesity
## 12           10 168.0000      74.00000            23    30.5     Obesity
## 13           10 139.0000      80.00000            23    30.5  Overweight
## 14            1 189.0000      60.00000            23   846.0     Obesity
## 15            5 166.0000      72.00000            19   175.0  Overweight
## 16            7 100.0000      69.10547            23    30.5     Obesity
## 17            0 118.0000      84.00000            47   230.0     Obesity
## 18            7 107.0000      74.00000            23    30.5  Overweight
## 19            1 103.0000      30.00000            38    83.0     Obesity
## 20            1 115.0000      70.00000            30    96.0     Obesity
## 21            3 126.0000      88.00000            41   235.0     Obesity
## 22            8  99.0000      84.00000            23    30.5     Obesity
## 23            7 196.0000      90.00000            23    30.5     Obesity
## 24            9 119.0000      80.00000            35    30.5  Overweight
## 25           11 143.0000      94.00000            33   146.0     Obesity
## 26           10 125.0000      70.00000            26   115.0     Obesity
## 27            7 147.0000      76.00000            23    30.5     Obesity
## 28            1  97.0000      66.00000            15   140.0      Normal
## 29           13 145.0000      82.00000            19   110.0      Normal
## 30            5 117.0000      92.00000            23    30.5     Obesity
## 31            5 109.0000      75.00000            26    30.5     Obesity
## 32            3 158.0000      76.00000            36   245.0     Obesity
## 33            3  88.0000      58.00000            11    54.0      Normal
## 34            6  92.0000      92.00000            23    30.5      Normal
## 35           10 122.0000      78.00000            31    30.5  Overweight
```

```
## 36             4 103.0000     60.00000            33  192.0     Normal
## 37            11 138.0000     76.00000            23   30.5    Obesity
## 38             9 102.0000     76.00000            37   30.5    Obesity
## 39             2  90.0000     68.00000            42   30.5    Obesity
## 40             4 111.0000     72.00000            47  207.0    Obesity
## 41             3 180.0000     64.00000            25   70.0    Obesity
## 42             7 133.0000     84.00000            23   30.5    Obesity
## 43             7 106.0000     92.00000            18   30.5     Normal
## 44             9 171.0000    110.00000            24  240.0    Obesity
## 45             7 159.0000     64.00000            23   30.5 Overweight
## 46             0 180.0000     66.00000            39   30.5    Obesity
## 47             1 146.0000     56.00000            23   30.5 Overweight
## 48             2  71.0000     70.00000            27   30.5 Overweight
## 49             7 103.0000     66.00000            32   30.5    Obesity
## 50             7 105.0000     69.10547            23   30.5    Obesity
## 51             1 103.0000     80.00000            11   82.0     Normal
## 52             1 101.0000     50.00000            15   36.0     Normal
## 53             5  88.0000     66.00000            21   23.0     Normal
## 54             8 176.0000     90.00000            34  300.0    Obesity
## 55             7 150.0000     66.00000            42  342.0    Obesity
## 56             1  73.0000     50.00000            10   30.5     Normal
## 57             7 187.0000     68.00000            39  304.0    Obesity
## 58             0 100.0000     88.00000            60  110.0    Obesity
## 59             0 146.0000     82.00000            23   30.5    Obesity
## 60             0 105.0000     64.00000            41  142.0    Obesity
## 61             2  84.0000     69.10547            23   30.5    Obesity
## 62             8 133.0000     72.00000            23   30.5    Obesity
## 63             5  44.0000     62.00000            23   30.5 Overweight
## 64             2 141.0000     58.00000            34  128.0 Overweight
## 65             7 114.0000     66.00000            23   30.5    Obesity
## 66             5  99.0000     74.00000            27   30.5 Overweight
## 67             0 109.0000     88.00000            30   30.5    Obesity
## 68             2 109.0000     92.00000            23   30.5    Obesity
## 69             1  95.0000     66.00000            13   38.0     Normal
## 70             4 146.0000     85.00000            27  100.0 Overweight
## 71             2 100.0000     66.00000            20   90.0    Obesity
## 72             5 139.0000     64.00000            35  140.0 Overweight
## 73            13 126.0000     90.00000            23   30.5    Obesity
## 74             4 129.0000     86.00000            20  270.0    Obesity
## 75             1  79.0000     75.00000            30   30.5    Obesity
## 76             1 120.8945     48.00000            20   30.5     Normal
## 77             7  62.0000     78.00000            23   30.5    Obesity
## 78             5  95.0000     72.00000            33   30.5    Obesity
## 79             0 131.0000     69.10547            23   30.5    Obesity
## 80             2 112.0000     66.00000            22   30.5 Overweight
## 81             3 113.0000     44.00000            13   30.5     Normal
## 82             2  74.0000     69.10547            23   30.5    Obesity
## 83             7  83.0000     78.00000            26   71.0 Overweight
## 84             0 101.0000     65.00000            28   30.5     Normal
## 85             5 137.0000    108.00000            23   30.5    Obesity
## 86             2 110.0000     74.00000            29  125.0    Obesity
## 87            13 106.0000     72.00000            54   30.5    Obesity
## 88             2 100.0000     68.00000            25   71.0    Obesity
## 89            15 136.0000     70.00000            32  110.0    Obesity
```

```
## 90             1 107.0000     68.00000              19    30.5 Overweight
## 91             1  80.0000     55.00000              23    30.5     Normal
## 92             4 123.0000     80.00000              15   176.0    Obesity
## 93             7  81.0000     78.00000              40    48.0    Obesity
## 94             4 134.0000     72.00000              23    30.5     Normal
## 95             2 142.0000     82.00000              18    64.0     Normal
## 96             6 144.0000     72.00000              27   228.0    Obesity
## 97             2  92.0000     62.00000              28    30.5    Obesity
## 98             1  71.0000     48.00000              18    76.0     Normal
## 99             6  93.0000     50.00000              30    64.0 Overweight
## 100            1 122.0000     90.00000              51   220.0    Obesity
## 101            1 163.0000     72.00000              23    30.5    Obesity
## 102            1 151.0000     60.00000              23    30.5 Overweight
## 103            0 125.0000     96.00000              23    30.5     Normal
## 104            1  81.0000     72.00000              18    40.0 Overweight
## 105            2  85.0000     65.00000              23    30.5    Obesity
## 106            1 126.0000     56.00000              29   152.0 Overweight
## 107            1  96.0000    122.00000              23    30.5     Normal
## 108            4 144.0000     58.00000              28   140.0 Overweight
## 109            3  83.0000     58.00000              31    18.0    Obesity
## 110            0  95.0000     85.00000              25    36.0    Obesity
## 111            3 171.0000     72.00000              33   135.0    Obesity
## 112            8 155.0000     62.00000              26   495.0    Obesity
## 113            1  89.0000     76.00000              34    37.0    Obesity
## 114            4  76.0000     62.00000              23    30.5    Obesity
## 115            7 160.0000     54.00000              32   175.0    Obesity
## 116            4 146.0000     92.00000              23    30.5    Obesity
## 117            5 124.0000     74.00000              23    30.5    Obesity
## 118            5  78.0000     48.00000              23    30.5    Obesity
## 119            4  97.0000     60.00000              23    30.5 Overweight
## 120            4  99.0000     76.00000              15    51.0     Normal
## 121            0 162.0000     76.00000              56   100.0    Obesity
## 122            6 111.0000     64.00000              39    30.5    Obesity
## 123            2 107.0000     74.00000              30   100.0    Obesity
## 124            5 132.0000     80.00000              23    30.5 Overweight
## 125            0 113.0000     76.00000              23    30.5    Obesity
## 126            1  88.0000     30.00000              42    99.0    Obesity
## 127            3 120.0000     70.00000              30   135.0    Obesity
## 128            1 118.0000     58.00000              36    94.0    Obesity
## 129            1 117.0000     88.00000              24   145.0    Obesity
## 130            0 105.0000     84.00000              23    30.5 Overweight
## 131            4 173.0000     70.00000              14   168.0 Overweight
## 132            9 122.0000     56.00000              23    30.5    Obesity
## 133            3 170.0000     64.00000              37   225.0    Obesity
## 134            8  84.0000     74.00000              31    30.5    Obesity
## 135            2  96.0000     68.00000              13    49.0     Normal
## 136            2 125.0000     60.00000              20   140.0    Obesity
## 137            0 100.0000     70.00000              26    50.0    Obesity
## 138            0  93.0000     60.00000              25    92.0 Overweight
## 139            0 129.0000     80.00000              23    30.5    Obesity
## 140            5 105.0000     72.00000              29   325.0    Obesity
## 141            3 128.0000     78.00000              23    30.5     Normal
## 142            5 106.0000     82.00000              30    30.5    Obesity
## 143            2 108.0000     52.00000              26    63.0    Obesity
```

```
## 144           10 108.0000     66.00000         23   30.5     Obesity
## 145            4 154.0000     62.00000         31  284.0     Obesity
## 146            0 102.0000     75.00000         23   30.5     Obesity
## 147            9  57.0000     80.00000         37   30.5     Obesity
## 148            2 106.0000     64.00000         35  119.0     Obesity
## 149            5 147.0000     78.00000         23   30.5     Obesity
## 150            2  90.0000     70.00000         17   30.5  Overweight
## 151            1 136.0000     74.00000         50  204.0     Obesity
## 152            4 114.0000     65.00000         23   30.5      Normal
## 153            9 156.0000     86.00000         28  155.0     Obesity
## 154            1 153.0000     82.00000         42  485.0     Obesity
## 155            8 188.0000     78.00000         23   30.5     Obesity
## 156            7 152.0000     88.00000         44   30.5     Obesity
## 157            2  99.0000     52.00000         15   94.0      Normal
## 158            1 109.0000     56.00000         21  135.0  Overweight
## 159            2  88.0000     74.00000         19   53.0  Overweight
## 160           17 163.0000     72.00000         41  114.0     Obesity
## 161            4 151.0000     90.00000         38   30.5  Overweight
## 162            7 102.0000     74.00000         40  105.0     Obesity
## 163            0 114.0000     80.00000         34  285.0     Obesity
## 164            2 100.0000     64.00000         23   30.5  Overweight
## 165            0 131.0000     88.00000         23   30.5     Obesity
## 166            6 104.0000     74.00000         18  156.0  Overweight
## 167            3 148.0000     66.00000         25   30.5     Obesity
## 168            4 120.0000     68.00000         23   30.5  Overweight
## 169            4 110.0000     66.00000         23   30.5     Obesity
## 170            3 111.0000     90.00000         12   78.0  Overweight
## 171            6 102.0000     82.00000         23   30.5     Obesity
## 172            6 134.0000     70.00000         23  130.0     Obesity
## 173            2  87.0000     69.10547         23   30.5  Overweight
## 174            1  79.0000     60.00000         42   48.0     Obesity
## 175            2  75.0000     64.00000         24   55.0  Overweight
## 176            8 179.0000     72.00000         42  130.0     Obesity
## 177            6  85.0000     78.00000         23   30.5     Obesity
## 178            0 129.0000    110.00000         46  130.0     Obesity
## 179            5 143.0000     78.00000         23   30.5     Obesity
## 180            5 130.0000     82.00000         23   30.5     Obesity
## 181            6  87.0000     80.00000         23   30.5      Normal
## 182            0 119.0000     64.00000         18   92.0     Obesity
## 183            1 120.8945     74.00000         20   23.0  Overweight
## 184            5  73.0000     60.00000         23   30.5  Overweight
## 185            4 141.0000     74.00000         23   30.5  Overweight
## 186            7 194.0000     68.00000         28   30.5     Obesity
## 187            8 181.0000     68.00000         36  495.0     Obesity
## 188            1 128.0000     98.00000         41   58.0     Obesity
## 189            8 109.0000     76.00000         39  114.0  Overweight
## 190            5 139.0000     80.00000         35  160.0     Obesity
## 191            3 111.0000     62.00000         23   30.5      Normal
## 192            9 123.0000     70.00000         44   94.0     Obesity
## 193            7 159.0000     66.00000         23   30.5     Obesity
## 194           11 135.0000     69.10547         23   30.5     Obesity
## 195            8  85.0000     55.00000         20   30.5      Normal
## 196            5 158.0000     84.00000         41  210.0     Obesity
## 197            1 105.0000     58.00000         23   30.5      Normal
```

```
## 198            3 107.0000       62.00000                13    48.0       Normal
## 199            4 109.0000       64.00000                44    99.0      Obesity
## 200            4 148.0000       60.00000                27   318.0      Obesity
## 201            0 113.0000       80.00000                16    30.5      Obesity
## 202            1 138.0000       82.00000                23    30.5      Obesity
## 203            0 108.0000       68.00000                20    30.5   Overweight
## 204            2  99.0000       70.00000                16    44.0       Normal
## 205            6 103.0000       72.00000                32   190.0      Obesity
## 206            5 111.0000       72.00000                28    30.5       Normal
## 207            8 196.0000       76.00000                29   280.0      Obesity
## 208            5 162.0000      104.00000                23    30.5      Obesity
## 209            1  96.0000       64.00000                27    87.0      Obesity
## 210            7 184.0000       84.00000                33    30.5      Obesity
## 211            2  81.0000       60.00000                22    30.5   Overweight
## 212            0 147.0000       85.00000                54    30.5      Obesity
## 213            7 179.0000       95.00000                31    30.5      Obesity
## 214            0 140.0000       65.00000                26   130.0      Obesity
## 215            9 112.0000       82.00000                32   175.0      Obesity
## 216           12 151.0000       70.00000                40   271.0      Obesity
## 217            5 109.0000       62.00000                41   129.0      Obesity
## 218            6 125.0000       68.00000                30   120.0      Obesity
## 219            5  85.0000       74.00000                22    30.5   Overweight
## 220            5 112.0000       66.00000                23    30.5      Obesity
## 221            0 177.0000       60.00000                29   478.0      Obesity
## 222            2 158.0000       90.00000                23    30.5      Obesity
## 223            7 119.0000       69.10547                23    30.5   Overweight
## 224            7 142.0000       60.00000                33   190.0   Overweight
## 225            1 100.0000       66.00000                15    56.0       Normal
## 226            1  87.0000       78.00000                27    32.0      Obesity
## 227            0 101.0000       76.00000                23    30.5      Obesity
## 228            3 162.0000       52.00000                38    30.5      Obesity
## 229            4 197.0000       70.00000                39   744.0      Obesity
## 230            0 117.0000       80.00000                31    53.0      Obesity
## 231            4 142.0000       86.00000                23    30.5      Obesity
## 232            6 134.0000       80.00000                37   370.0      Obesity
## 233            1  79.0000       80.00000                25    37.0   Overweight
## 234            4 122.0000       68.00000                23    30.5      Obesity
## 235            3  74.0000       68.00000                28    45.0   Overweight
## 236            4 171.0000       72.00000                23    30.5      Obesity
## 237            7 181.0000       84.00000                21   192.0      Obesity
## 238            0 179.0000       90.00000                27    30.5      Obesity
## 239            9 164.0000       84.00000                21    30.5      Obesity
## 240            0 104.0000       76.00000                23    30.5  Underweight
## 241            1  91.0000       64.00000                24    30.5   Overweight
## 242            4  91.0000       70.00000                32    88.0      Obesity
## 243            3 139.0000       54.00000                23    30.5   Overweight
## 244            6 119.0000       50.00000                22   176.0   Overweight
## 245            2 146.0000       76.00000                35   194.0      Obesity
## 246            9 184.0000       85.00000                15    30.5      Obesity
## 247           10 122.0000       68.00000                23    30.5      Obesity
## 248            0 165.0000       90.00000                33   680.0      Obesity
## 249            9 124.0000       70.00000                33   402.0      Obesity
## 250            1 111.0000       86.00000                19    30.5      Obesity
## 251            9 106.0000       52.00000                23    30.5      Obesity
```

```
## 252            2 129.0000       84.00000             23    30.5 Overweight
## 253            2  90.0000       80.00000             14    55.0     Normal
## 254            0  86.0000       68.00000             32    30.5    Obesity
## 255           12  92.0000       62.00000              7   258.0 Overweight
## 256            1 113.0000       64.00000             35    30.5    Obesity
## 257            3 111.0000       56.00000             39    30.5    Obesity
## 258            2 114.0000       68.00000             22    30.5 Overweight
## 259            1 193.0000       50.00000             16   375.0 Overweight
## 260           11 155.0000       76.00000             28   150.0    Obesity
## 261            3 191.0000       68.00000             15   130.0    Obesity
## 262            3 141.0000       69.10547             23    30.5    Obesity
## 263            4  95.0000       70.00000             32    30.5    Obesity
## 264            3 142.0000       80.00000             15    30.5    Obesity
## 265            4 123.0000       62.00000             23    30.5    Obesity
## 266            5  96.0000       74.00000             18    67.0    Obesity
## 267            0 138.0000       69.10547             23    30.5    Obesity
## 268            2 128.0000       64.00000             42    30.5    Obesity
## 269            0 102.0000       52.00000             23    30.5 Overweight
## 270            2 146.0000       69.10547             23    30.5 Overweight
## 271           10 101.0000       86.00000             37    30.5    Obesity
## 272            2 108.0000       62.00000             32    56.0 Overweight
## 273            3 122.0000       78.00000             23    30.5     Normal
## 274            1  71.0000       78.00000             50    45.0    Obesity
## 275           13 106.0000       70.00000             23    30.5    Obesity
## 276            2 100.0000       70.00000             52    57.0    Obesity
## 277            7 106.0000       60.00000             24    30.5 Overweight
## 278            0 104.0000       64.00000             23   116.0 Overweight
## 279            5 114.0000       74.00000             23    30.5     Normal
## 280            2 108.0000       62.00000             10   278.0 Overweight
## 281            0 146.0000       70.00000             23    30.5    Obesity
## 282           10 129.0000       76.00000             28   122.0    Obesity
## 283            7 133.0000       88.00000             15   155.0    Obesity
## 284            7 161.0000       86.00000             23    30.5    Obesity
## 285            2 108.0000       80.00000             23    30.5 Overweight
## 286            7 136.0000       74.00000             26   135.0 Overweight
## 287            5 155.0000       84.00000             44   545.0    Obesity
## 288            1 119.0000       86.00000             39   220.0    Obesity
## 289            4  96.0000       56.00000             17    49.0     Normal
## 290            5 108.0000       72.00000             43    75.0    Obesity
## 291            0  78.0000       88.00000             29    40.0    Obesity
## 292            0 107.0000       62.00000             30    74.0    Obesity
## 293            2 128.0000       78.00000             37   182.0    Obesity
## 294            1 128.0000       48.00000             45   194.0    Obesity
## 295            0 161.0000       50.00000             23    30.5     Normal
## 296            6 151.0000       62.00000             31   120.0    Obesity
## 297            2 146.0000       70.00000             38   360.0 Overweight
## 298            0 126.0000       84.00000             29   215.0    Obesity
## 299           14 100.0000       78.00000             25   184.0    Obesity
## 300            8 112.0000       72.00000             23    30.5     Normal
## 301            0 167.0000       69.10547             23    30.5    Obesity
## 302            2 144.0000       58.00000             33   135.0    Obesity
## 303            5  77.0000       82.00000             41    42.0    Obesity
## 304            5 115.0000       98.00000             23    30.5    Obesity
## 305            3 150.0000       76.00000             23    30.5     Normal
```

```
## 306            2 120.0000       76.00000            37   105.0     Obesity
## 307           10 161.0000       68.00000            23   132.0  Overweight
## 308            0 137.0000       68.00000            14   148.0      Normal
## 309            0 128.0000       68.00000            19   180.0     Obesity
## 310            2 124.0000       68.00000            28   205.0     Obesity
## 311            6  80.0000       66.00000            30    30.5  Overweight
## 312            0 106.0000       70.00000            37   148.0     Obesity
## 313            2 155.0000       74.00000            17    96.0  Overweight
## 314            3 113.0000       50.00000            10    85.0  Overweight
## 315            7 109.0000       80.00000            31    30.5     Obesity
## 316            2 112.0000       68.00000            22    94.0     Obesity
## 317            3  99.0000       80.00000            11    64.0      Normal
## 318            3 182.0000       74.00000            23    30.5     Obesity
## 319            3 115.0000       66.00000            39   140.0     Obesity
## 320            6 194.0000       78.00000            23    30.5      Normal
## 321            4 129.0000       60.00000            12   231.0  Overweight
## 322            3 112.0000       74.00000            30    30.5     Obesity
## 323            0 124.0000       70.00000            20    30.5  Overweight
## 324           13 152.0000       90.00000            33    29.0  Overweight
## 325            2 112.0000       75.00000            32    30.5     Obesity
## 326            1 157.0000       72.00000            21   168.0  Overweight
## 327            1 122.0000       64.00000            32   156.0     Obesity
## 328           10 179.0000       70.00000            23    30.5     Obesity
## 329            2 102.0000       86.00000            36   120.0     Obesity
## 330            6 105.0000       70.00000            32    68.0     Obesity
## 331            8 118.0000       72.00000            19    30.5      Normal
## 332            2  87.0000       58.00000            16    52.0     Obesity
## 333            1 180.0000       69.10547            23    30.5     Obesity
## 334           12 106.0000       80.00000            23    30.5      Normal
## 335            1  95.0000       60.00000            18    58.0      Normal
## 336            0 165.0000       76.00000            43   255.0     Obesity
## 337            0 117.0000       69.10547            23    30.5     Obesity
## 338            5 115.0000       76.00000            23    30.5     Obesity
## 339            9 152.0000       78.00000            34   171.0     Obesity
## 340            7 178.0000       84.00000            23    30.5     Obesity
## 341            1 130.0000       70.00000            13   105.0  Overweight
## 342            1  95.0000       74.00000            21    73.0  Overweight
## 343            1 120.8945       68.00000            35    30.5     Obesity
## 344            5 122.0000       86.00000            23    30.5     Obesity
## 345            8  95.0000       72.00000            23    30.5     Obesity
## 346            8 126.0000       88.00000            36   108.0     Obesity
## 347            1 139.0000       46.00000            19    83.0  Overweight
## 348            3 116.0000       69.10547            23    30.5      Normal
## 349            3  99.0000       62.00000            19    74.0      Normal
## 350            5 120.8945       80.00000            32    30.5     Obesity
## 351            4  92.0000       80.00000            23    30.5     Obesity
## 352            4 137.0000       84.00000            23    30.5     Obesity
## 353            3  61.0000       82.00000            28    30.5     Obesity
## 354            1  90.0000       62.00000            12    43.0  Overweight
## 355            3  90.0000       78.00000            23    30.5     Obesity
## 356            9 165.0000       88.00000            23    30.5     Obesity
## 357            1 125.0000       50.00000            40   167.0     Obesity
## 358           13 129.0000       69.10547            30    30.5     Obesity
## 359           12  88.0000       74.00000            40    54.0     Obesity
```

```
## 360             1 196.0000        76.00000              36   249.0     Obesity
## 361             5 189.0000        64.00000              33   325.0     Obesity
## 362             5 158.0000        70.00000              23    30.5  Overweight
## 363             5 103.0000       108.00000              37    30.5     Obesity
## 364             4 146.0000        78.00000              23    30.5     Obesity
## 365             4 147.0000        74.00000              25   293.0     Obesity
## 366             5  99.0000        54.00000              28    83.0     Obesity
## 367             6 124.0000        72.00000              23    30.5  Overweight
## 368             0 101.0000        64.00000              17    30.5      Normal
## 369             3  81.0000        86.00000              16    66.0  Overweight
## 370             1 133.0000       102.00000              28   140.0     Obesity
## 371             3 173.0000        82.00000              48   465.0     Obesity
## 372             0 118.0000        64.00000              23    89.0     Obesity
## 373             0  84.0000        64.00000              22    66.0     Obesity
## 374             2 105.0000        58.00000              40    94.0     Obesity
## 375             2 122.0000        52.00000              43   158.0     Obesity
## 376            12 140.0000        82.00000              43   325.0     Obesity
## 377             0  98.0000        82.00000              15    84.0  Overweight
## 378             1  87.0000        60.00000              37    75.0     Obesity
## 379             4 156.0000        75.00000              23    30.5     Obesity
## 380             0  93.0000       100.00000              39    72.0     Obesity
## 381             1 107.0000        72.00000              30    82.0     Obesity
## 382             0 105.0000        68.00000              22    30.5      Normal
## 383             1 109.0000        60.00000               8   182.0  Overweight
## 384             1  90.0000        62.00000              18    59.0  Overweight
## 385             1 125.0000        70.00000              24   110.0      Normal
## 386             1 119.0000        54.00000              13    50.0      Normal
## 387             5 116.0000        74.00000              29    30.5     Obesity
## 388             8 105.0000       100.00000              36    30.5     Obesity
## 389             5 144.0000        82.00000              26   285.0     Obesity
## 390             3 100.0000        68.00000              23    81.0     Obesity
## 391             1 100.0000        66.00000              29   196.0     Obesity
## 392             5 166.0000        76.00000              23    30.5     Obesity
## 393             1 131.0000        64.00000              14   415.0      Normal
## 394             4 116.0000        72.00000              12    87.0      Normal
## 395             4 158.0000        78.00000              23    30.5     Obesity
## 396             2 127.0000        58.00000              24   275.0  Overweight
## 397             3  96.0000        56.00000              34   115.0      Normal
## 398             0 131.0000        66.00000              40    30.5     Obesity
## 399             3  82.0000        70.00000              23    30.5      Normal
## 400             3 193.0000        70.00000              31    30.5     Obesity
## 401             4  95.0000        64.00000              23    30.5     Obesity
## 402             6 137.0000        61.00000              23    30.5      Normal
## 403             5 136.0000        84.00000              41    88.0     Obesity
## 404             9  72.0000        78.00000              25    30.5     Obesity
## 405             5 168.0000        64.00000              23    30.5     Obesity
## 406             2 123.0000        48.00000              32   165.0     Obesity
## 407             4 115.0000        72.00000              23    30.5  Overweight
## 408             0 101.0000        62.00000              23    30.5      Normal
## 409             8 197.0000        74.00000              23    30.5  Overweight
## 410             1 172.0000        68.00000              49   579.0     Obesity
## 411             6 102.0000        90.00000              39    30.5     Obesity
## 412             1 112.0000        72.00000              30   176.0     Obesity
## 413             1 143.0000        84.00000              23   310.0     Obesity
```

```
## 414             1 143.0000       74.00000             22    61.0  Overweight
## 415             0 138.0000       60.00000             35   167.0     Obesity
## 416             3 173.0000       84.00000             33   474.0     Obesity
## 417             1  97.0000       68.00000             21    30.5  Overweight
## 418             4 144.0000       82.00000             32    30.5     Obesity
## 419             1  83.0000       68.00000             23    30.5 Underweight
## 420             3 129.0000       64.00000             29   115.0  Overweight
## 421             1 119.0000       88.00000             41   170.0     Obesity
## 422             2  94.0000       68.00000             18    76.0  Overweight
## 423             0 102.0000       64.00000             46    78.0     Obesity
## 424             2 115.0000       64.00000             22    30.5     Obesity
## 425             8 151.0000       78.00000             32   210.0     Obesity
## 426             4 184.0000       78.00000             39   277.0     Obesity
## 427             0  94.0000       69.10547             23    30.5     Obesity
## 428             1 181.0000       64.00000             30   180.0     Obesity
## 429             0 135.0000       94.00000             46   145.0     Obesity
## 430             1  95.0000       82.00000             25   180.0     Obesity
## 431             2  99.0000       69.10547             23    30.5      Normal
## 432             3  89.0000       74.00000             16    85.0     Obesity
## 433             1  80.0000       74.00000             11    60.0     Obesity
## 434             2 139.0000       75.00000             23    30.5  Overweight
## 435             1  90.0000       68.00000              8    30.5      Normal
## 436             0 141.0000       69.10547             23    30.5     Obesity
## 437            12 140.0000       85.00000             33    30.5     Obesity
## 438             5 147.0000       75.00000             23    30.5  Overweight
## 439             1  97.0000       70.00000             15    30.5 Underweight
## 440             6 107.0000       88.00000             23    30.5     Obesity
## 441             0 189.0000      104.00000             25    30.5     Obesity
## 442             2  83.0000       66.00000             23    50.0     Obesity
## 443             4 117.0000       64.00000             27   120.0     Obesity
## 444             8 108.0000       70.00000             23    30.5     Obesity
## 445             4 117.0000       62.00000             12    30.5  Overweight
## 446             0 180.0000       78.00000             63    14.0     Obesity
## 447             1 100.0000       72.00000             12    70.0  Overweight
## 448             0  95.0000       80.00000             45    92.0     Obesity
## 449             0 104.0000       64.00000             37    64.0     Obesity
## 450             0 120.0000       74.00000             18    63.0     Obesity
## 451             1  82.0000       64.00000             13    95.0      Normal
## 452             2 134.0000       70.00000             23    30.5  Overweight
## 453             0  91.0000       68.00000             32   210.0     Obesity
## 454             2 119.0000       69.10547             23    30.5      Normal
## 455             2 100.0000       54.00000             28   105.0     Obesity
## 456            14 175.0000       62.00000             30    30.5     Obesity
## 457             1 135.0000       54.00000             23    30.5  Overweight
## 458             5  86.0000       68.00000             28    71.0     Obesity
## 459            10 148.0000       84.00000             48   237.0     Obesity
## 460             9 134.0000       74.00000             33    60.0  Overweight
## 461             9 120.0000       72.00000             22    56.0      Normal
## 462             1  71.0000       62.00000             23    30.5      Normal
## 463             8  74.0000       70.00000             40    49.0     Obesity
## 464             5  88.0000       78.00000             30    30.5  Overweight
## 465            10 115.0000       98.00000             23    30.5      Normal
## 466             0 124.0000       56.00000             13   105.0      Normal
## 467             0  74.0000       52.00000             10    36.0  Overweight
```

```
## 468          0  97.0000      64.00000         36  100.0    Obesity
## 469          8 120.0000      69.10547         23   30.5    Obesity
## 470          6 154.0000      78.00000         41  140.0    Obesity
## 471          1 144.0000      82.00000         40   30.5    Obesity
## 472          0 137.0000      70.00000         38   30.5    Obesity
## 473          0 119.0000      66.00000         27   30.5    Obesity
## 474          7 136.0000      90.00000         23   30.5 Overweight
## 475          4 114.0000      64.00000         23   30.5 Overweight
## 476          0 137.0000      84.00000         27   30.5 Overweight
## 477          2 105.0000      80.00000         45  191.0    Obesity
## 478          7 114.0000      76.00000         17  110.0     Normal
## 479          8 126.0000      74.00000         38   75.0 Overweight
## 480          4 132.0000      86.00000         31   30.5 Overweight
## 481          3 158.0000      70.00000         30  328.0    Obesity
## 482          0 123.0000      88.00000         37   30.5    Obesity
## 483          4  85.0000      58.00000         22   49.0 Overweight
## 484          0  84.0000      82.00000         31  125.0    Obesity
## 485          0 145.0000      69.10547         23   30.5    Obesity
## 486          0 135.0000      68.00000         42  250.0    Obesity
## 487          1 139.0000      62.00000         41  480.0    Obesity
## 488          0 173.0000      78.00000         32  265.0    Obesity
## 489          4  99.0000      72.00000         17   30.5 Overweight
## 490          8 194.0000      80.00000         23   30.5 Overweight
## 491          2  83.0000      65.00000         28   66.0    Obesity
## 492          2  89.0000      90.00000         30   30.5    Obesity
## 493          4  99.0000      68.00000         38   30.5    Obesity
## 494          4 125.0000      70.00000         18  122.0 Overweight
## 495          3  80.0000      69.10547         23   30.5    Obesity
## 496          6 166.0000      74.00000         23   30.5 Overweight
## 497          5 110.0000      68.00000         23   30.5 Overweight
## 498          2  81.0000      72.00000         15   76.0    Obesity
## 499          7 195.0000      70.00000         33  145.0 Overweight
## 500          6 154.0000      74.00000         32  193.0 Overweight
## 501          2 117.0000      90.00000         19   71.0 Overweight
## 502          3  84.0000      72.00000         32   30.5    Obesity
## 503          6 120.8945      68.00000         41   30.5    Obesity
## 504          7  94.0000      64.00000         25   79.0    Obesity
## 505          3  96.0000      78.00000         39   30.5    Obesity
## 506         10  75.0000      82.00000         23   30.5    Obesity
## 507          0 180.0000      90.00000         26   90.0    Obesity
## 508          1 130.0000      60.00000         23  170.0 Overweight
## 509          2  84.0000      50.00000         23   76.0    Obesity
## 510          8 120.0000      78.00000         23   30.5 Overweight
## 511         12  84.0000      72.00000         31   30.5 Overweight
## 512          0 139.0000      62.00000         17  210.0     Normal
## 513          9  91.0000      68.00000         23   30.5     Normal
## 514          2  91.0000      62.00000         23   30.5 Overweight
## 515          3  99.0000      54.00000         19   86.0 Overweight
## 516          3 163.0000      70.00000         18  105.0    Obesity
## 517          9 145.0000      88.00000         34  165.0    Obesity
## 518          7 125.0000      86.00000         23   30.5    Obesity
## 519         13  76.0000      60.00000         23   30.5    Obesity
## 520          6 129.0000      90.00000          7  326.0     Normal
## 521          2  68.0000      70.00000         32   66.0 Overweight
```

```
## 522           3 124.0000       80.00000             33   130.0     Obesity
## 523           6 114.0000       69.10547             23    30.5     Obesity
## 524           9 130.0000       70.00000             23    30.5     Obesity
## 525           3 125.0000       58.00000             23    30.5     Obesity
## 526           3  87.0000       60.00000             18    30.5      Normal
## 527           1  97.0000       64.00000             19    82.0 Underweight
## 528           3 116.0000       74.00000             15   105.0  Overweight
## 529           0 117.0000       66.00000             31   188.0     Obesity
## 530           0 111.0000       65.00000             23    30.5      Normal
## 531           2 122.0000       60.00000             18   106.0  Overweight
## 532           0 107.0000       76.00000             23    30.5     Obesity
## 533           1  86.0000       66.00000             52    65.0     Obesity
## 534           6  91.0000       69.10547             23    30.5  Overweight
## 535           1  77.0000       56.00000             30    56.0     Obesity
## 536           4 132.0000       69.10547             23    30.5     Obesity
## 537           0 105.0000       90.00000             23    30.5  Overweight
## 538           0  57.0000       60.00000             23    30.5      Normal
## 539           0 127.0000       80.00000             37   210.0     Obesity
## 540           3 129.0000       92.00000             49   155.0     Obesity
## 541           8 100.0000       74.00000             40   215.0     Obesity
## 542           3 128.0000       72.00000             25   190.0     Obesity
## 543          10  90.0000       85.00000             32    30.5     Obesity
## 544           4  84.0000       90.00000             23    56.0     Obesity
## 545           1  88.0000       78.00000             29    76.0     Obesity
## 546           8 186.0000       90.00000             35   225.0     Obesity
## 547           5 187.0000       76.00000             27   207.0     Obesity
## 548           4 131.0000       68.00000             21   166.0     Obesity
## 549           1 164.0000       82.00000             43    67.0     Obesity
## 550           4 189.0000      110.00000             31    30.5  Overweight
## 551           1 116.0000       70.00000             28    30.5  Overweight
## 552           3  84.0000       68.00000             30   106.0     Obesity
## 553           6 114.0000       88.00000             23    30.5  Overweight
## 554           1  88.0000       62.00000             24    44.0  Overweight
## 555           1  84.0000       64.00000             23   115.0     Obesity
## 556           7 124.0000       70.00000             33   215.0  Overweight
## 557           1  97.0000       70.00000             40    30.5     Obesity
## 558           8 110.0000       76.00000             23    30.5  Overweight
## 559          11 103.0000       68.00000             40    30.5     Obesity
## 560          11  85.0000       74.00000             23    30.5     Obesity
## 561           6 125.0000       76.00000             23    30.5     Obesity
## 562           0 198.0000       66.00000             32   274.0     Obesity
## 563           1  87.0000       68.00000             34    77.0     Obesity
## 564           6  99.0000       60.00000             19    54.0  Overweight
## 565           0  91.0000       80.00000             23    30.5     Obesity
## 566           2  95.0000       54.00000             14    88.0  Overweight
## 567           1  99.0000       72.00000             30    18.0     Obesity
## 568           6  92.0000       62.00000             32   126.0     Obesity
## 569           4 154.0000       72.00000             29   126.0     Obesity
## 570           0 121.0000       66.00000             30   165.0     Obesity
## 571           3  78.0000       70.00000             23    30.5     Obesity
## 572           2 130.0000       96.00000             23    30.5      Normal
## 573           3 111.0000       58.00000             31    44.0  Overweight
## 574           2  98.0000       60.00000             17   120.0     Obesity
## 575           1 143.0000       86.00000             30   330.0     Obesity
```

```
## 576                1 119.0000       44.00000              47    63.0     Obesity
## 577                6 108.0000       44.00000              20   130.0      Normal
## 578                2 118.0000       80.00000              23    30.5     Obesity
## 579               10 133.0000       68.00000              23    30.5  Overweight
## 580                2 197.0000       70.00000              99    30.5     Obesity
## 581                0 151.0000       90.00000              46    30.5     Obesity
## 582                6 109.0000       60.00000              27    30.5  Overweight
## 583               12 121.0000       78.00000              17    30.5  Overweight
## 584                8 100.0000       76.00000              23    30.5     Obesity
## 585                8 124.0000       76.00000              24   600.0  Overweight
## 586                1  93.0000       56.00000              11    30.5      Normal
## 587                8 143.0000       66.00000              23    30.5     Obesity
## 588                6 103.0000       66.00000              23    30.5      Normal
## 589                3 176.0000       86.00000              27   156.0     Obesity
## 590                0  73.0000       69.10547              23    30.5      Normal
## 591               11 111.0000       84.00000              40    30.5     Obesity
## 592                2 112.0000       78.00000              50   140.0     Obesity
## 593                3 132.0000       80.00000              23    30.5     Obesity
## 594                2  82.0000       52.00000              22   115.0  Overweight
## 595                6 123.0000       72.00000              45   230.0     Obesity
## 596                0 188.0000       82.00000              14   185.0     Obesity
## 597                0  67.0000       76.00000              23    30.5     Obesity
## 598                1  89.0000       24.00000              19    25.0  Overweight
## 599                1 173.0000       74.00000              23    30.5     Obesity
## 600                1 109.0000       38.00000              18   120.0      Normal
## 601                1 108.0000       88.00000              19    30.5  Overweight
## 602                6  96.0000       69.10547              23    30.5      Normal
## 603                1 124.0000       74.00000              36    30.5  Overweight
## 604                7 150.0000       78.00000              29   126.0     Obesity
## 605                4 183.0000       69.10547              23    30.5  Overweight
## 606                1 124.0000       60.00000              32    30.5     Obesity
## 607                1 181.0000       78.00000              42   293.0     Obesity
## 608                1  92.0000       62.00000              25    41.0      Normal
## 609                0 152.0000       82.00000              39   272.0     Obesity
## 610                1 111.0000       62.00000              13   182.0      Normal
## 611                3 106.0000       54.00000              21   158.0     Obesity
## 612                3 174.0000       58.00000              22   194.0     Obesity
## 613                7 168.0000       88.00000              42   321.0     Obesity
## 614                6 105.0000       80.00000              28    30.5     Obesity
## 615               11 138.0000       74.00000              26   144.0     Obesity
## 616                3 106.0000       72.00000              23    30.5  Overweight
## 617                6 117.0000       96.00000              23    30.5  Overweight
## 618                2  68.0000       62.00000              13    15.0      Normal
## 619                9 112.0000       82.00000              24    30.5  Overweight
## 620                0 119.0000       69.10547              23    30.5     Obesity
## 621                2 112.0000       86.00000              42   160.0     Obesity
## 622                2  92.0000       76.00000              20    30.5      Normal
## 623                6 183.0000       94.00000              23    30.5     Obesity
## 624                0  94.0000       70.00000              27   115.0     Obesity
## 625                2 108.0000       64.00000              23    30.5     Obesity
## 626                4  90.0000       88.00000              47    54.0     Obesity
## 627                0 125.0000       68.00000              23    30.5      Normal
## 628                0 132.0000       78.00000              23    30.5     Obesity
## 629                5 128.0000       80.00000              23    30.5     Obesity
```

```
## 630           4  94.0000       65.00000             22    30.5      Normal
## 631           7 114.0000       64.00000             23    30.5  Overweight
## 632           0 102.0000       78.00000             40    90.0     Obesity
## 633           2 111.0000       60.00000             23    30.5  Overweight
## 634           1 128.0000       82.00000             17   183.0  Overweight
## 635          10  92.0000       62.00000             23    30.5  Overweight
## 636          13 104.0000       72.00000             23    30.5     Obesity
## 637           5 104.0000       74.00000             23    30.5  Overweight
## 638           2  94.0000       76.00000             18    66.0     Obesity
## 639           7  97.0000       76.00000             32    91.0     Obesity
## 640           1 100.0000       74.00000             12    46.0      Normal
## 641           0 102.0000       86.00000             17   105.0  Overweight
## 642           4 128.0000       70.00000             23    30.5     Obesity
## 643           6 147.0000       80.00000             23    30.5  Overweight
## 644           4  90.0000       69.10547             23    30.5  Overweight
## 645           3 103.0000       72.00000             30   152.0  Overweight
## 646           2 157.0000       74.00000             35   440.0     Obesity
## 647           1 167.0000       74.00000             17   144.0      Normal
## 648           0 179.0000       50.00000             36   159.0     Obesity
## 649          11 136.0000       84.00000             35   130.0  Overweight
## 650           0 107.0000       60.00000             25    30.5  Overweight
## 651           1  91.0000       54.00000             25   100.0  Overweight
## 652           1 117.0000       60.00000             23   106.0     Obesity
## 653           5 123.0000       74.00000             40    77.0     Obesity
## 654           2 120.0000       54.00000             23    30.5  Overweight
## 655           1 106.0000       70.00000             28   135.0     Obesity
## 656           2 155.0000       52.00000             27   540.0     Obesity
## 657           2 101.0000       58.00000             35    90.0      Normal
## 658           1 120.0000       80.00000             48   200.0     Obesity
## 659          11 127.0000      106.00000             23    30.5     Obesity
## 660           3  80.0000       82.00000             31    70.0     Obesity
## 661          10 162.0000       84.00000             23    30.5  Overweight
## 662           1 199.0000       76.00000             43    30.5     Obesity
## 663           8 167.0000      106.00000             46   231.0     Obesity
## 664           9 145.0000       80.00000             46   130.0     Obesity
## 665           6 115.0000       60.00000             39    30.5     Obesity
## 666           1 112.0000       80.00000             45   132.0     Obesity
## 667           4 145.0000       82.00000             18    30.5     Obesity
## 668          10 111.0000       70.00000             27    30.5  Overweight
## 669           6  98.0000       58.00000             33   190.0     Obesity
## 670           9 154.0000       78.00000             30   100.0     Obesity
## 671           6 165.0000       68.00000             26   168.0     Obesity
## 672           1  99.0000       58.00000             10    30.5  Overweight
## 673          10  68.0000      106.00000             23    49.0     Obesity
## 674           3 123.0000      100.00000             35   240.0     Obesity
## 675           8  91.0000       82.00000             23    30.5     Obesity
## 676           6 195.0000       70.00000             23    30.5     Obesity
## 677           9 156.0000       86.00000             23    30.5      Normal
## 678           0  93.0000       60.00000             23    30.5     Obesity
## 679           3 121.0000       52.00000             23    30.5     Obesity
## 680           2 101.0000       58.00000             17   265.0      Normal
## 681           2  56.0000       56.00000             28    45.0      Normal
## 682           0 162.0000       76.00000             36    30.5     Obesity
## 683           0  95.0000       64.00000             39   105.0     Obesity
```

```
## 684               4 125.0000       80.00000             23    30.5     Obesity
## 685               5 136.0000       82.00000             23    30.5     Obesity
## 686               2 129.0000       74.00000             26   205.0     Obesity
## 687               3 130.0000       64.00000             23    30.5      Normal
## 688               1 107.0000       50.00000             19    30.5  Overweight
## 689               1 140.0000       74.00000             26   180.0      Normal
## 690               1 144.0000       82.00000             46   180.0     Obesity
## 691               8 107.0000       80.00000             23    30.5      Normal
## 692              13 158.0000      114.00000             23    30.5     Obesity
## 693               2 121.0000       70.00000             32    95.0     Obesity
## 694               7 129.0000       68.00000             49   125.0     Obesity
## 695               2  90.0000       60.00000             23    30.5      Normal
## 696               7 142.0000       90.00000             24   480.0     Obesity
## 697               3 169.0000       74.00000             19   125.0  Overweight
## 698               0  99.0000       69.10547             23    30.5  Overweight
## 699               4 127.0000       88.00000             11   155.0     Obesity
## 700               4 118.0000       70.00000             23    30.5     Obesity
## 701               2 122.0000       76.00000             27   200.0     Obesity
## 702               6 125.0000       78.00000             31    30.5  Overweight
## 703               1 168.0000       88.00000             29    30.5     Obesity
## 704               2 129.0000       69.10547             23    30.5     Obesity
## 705               4 110.0000       76.00000             20   100.0  Overweight
## 706               6  80.0000       80.00000             36    30.5     Obesity
## 707              10 115.0000       69.10547             23    30.5     Obesity
## 708               2 127.0000       46.00000             21   335.0     Obesity
## 709               9 164.0000       78.00000             23    30.5     Obesity
## 710               2  93.0000       64.00000             32   160.0     Obesity
## 711               3 158.0000       64.00000             13   387.0     Obesity
## 712               5 126.0000       78.00000             27    22.0  Overweight
## 713              10 129.0000       62.00000             36    30.5     Obesity
## 714               0 134.0000       58.00000             20   291.0  Overweight
## 715               3 102.0000       74.00000             23    30.5  Overweight
## 716               7 187.0000       50.00000             33   392.0     Obesity
## 717               3 173.0000       78.00000             39   185.0     Obesity
## 718              10  94.0000       72.00000             18    30.5      Normal
## 719               1 108.0000       60.00000             46   178.0     Obesity
## 720               5  97.0000       76.00000             27    30.5     Obesity
## 721               4  83.0000       86.00000             19    30.5  Overweight
## 722               1 114.0000       66.00000             36   200.0     Obesity
## 723               1 149.0000       68.00000             29   127.0  Overweight
## 724               5 117.0000       86.00000             30   105.0     Obesity
## 725               1 111.0000       94.00000             23    30.5     Obesity
## 726               4 112.0000       78.00000             40    30.5     Obesity
## 727               1 116.0000       78.00000             29   180.0     Obesity
## 728               0 141.0000       84.00000             26    30.5     Obesity
## 729               2 175.0000       88.00000             23    30.5      Normal
## 730               2  92.0000       52.00000             23    30.5     Obesity
## 731               3 130.0000       78.00000             23    79.0  Overweight
## 732               8 120.0000       86.00000             23    30.5  Overweight
## 733               2 174.0000       88.00000             37   120.0     Obesity
## 734               2 106.0000       56.00000             27   165.0  Overweight
## 735               2 105.0000       75.00000             23    30.5      Normal
## 736               4  95.0000       60.00000             32    30.5     Obesity
## 737               0 126.0000       86.00000             27   120.0  Overweight
```

```
## 738           8  65.0000       72.00000             23    30.5     Obesity
## 739           2  99.0000       60.00000             17   160.0     Obesity
## 740           1 102.0000       74.00000             23    30.5     Obesity
## 741          11 120.0000       80.00000             37   150.0     Obesity
## 742           3 102.0000       44.00000             20    94.0     Obesity
## 743           1 109.0000       58.00000             18   116.0  Overweight
## 744           9 140.0000       94.00000             23    30.5     Obesity
## 745          13 153.0000       88.00000             37   140.0     Obesity
## 746          12 100.0000       84.00000             33   105.0     Obesity
## 747           1 147.0000       94.00000             41    30.5     Obesity
## 748           1  81.0000       74.00000             41    57.0     Obesity
## 749           3 187.0000       70.00000             22   200.0     Obesity
## 750           6 162.0000       62.00000             23    30.5      Normal
## 751           4 136.0000       70.00000             23    30.5     Obesity
## 752           1 121.0000       78.00000             39    74.0     Obesity
## 753           3 108.0000       62.00000             24    30.5  Overweight
## 754           0 181.0000       88.00000             44   510.0     Obesity
## 755           8 154.0000       78.00000             32    30.5     Obesity
## 756           1 128.0000       88.00000             39   110.0     Obesity
## 757           7 137.0000       90.00000             41    30.5     Obesity
## 758           0 123.0000       72.00000             23    30.5     Obesity
## 759           1 106.0000       76.00000             23    30.5     Obesity
## 760           6 190.0000       92.00000             23    30.5     Obesity
## 761           2  88.0000       58.00000             26    16.0  Overweight
## 762           9 170.0000       74.00000             31    30.5     Obesity
## 763           9  89.0000       62.00000             23    30.5      Normal
## 764          10 101.0000       76.00000             48   180.0     Obesity
## 765           2 122.0000       70.00000             27    30.5     Obesity
## 766           5 121.0000       72.00000             23   112.0  Overweight
## 767           1 126.0000       60.00000             23    30.5     Obesity
## 768           1  93.0000       70.00000             31    30.5     Obesity
##     DiabetesPedigreeFunction Age Outcome
## 1                      0.627  50       1
## 2                      0.351  31       0
## 3                      0.672  32       1
## 4                      0.167  21       0
## 5                      2.288  33       1
## 6                      0.201  30       0
## 7                      0.248  26       1
## 8                      0.134  29       0
## 9                      0.158  53       1
## 10                     0.232  54       1
## 11                     0.191  30       0
## 12                     0.537  34       1
## 13                     1.441  57       0
## 14                     0.398  59       1
## 15                     0.587  51       1
## 16                     0.484  32       1
## 17                     0.551  31       1
## 18                     0.254  31       1
## 19                     0.183  33       0
## 20                     0.529  32       1
## 21                     0.704  27       0
## 22                     0.388  50       0
```

```
## 23                        0.451  41         1
## 24                        0.263  29         1
## 25                        0.254  51         1
## 26                        0.205  41         1
## 27                        0.257  43         1
## 28                        0.487  22         0
## 29                        0.245  57         0
## 30                        0.337  38         0
## 31                        0.546  60         0
## 32                        0.851  28         1
## 33                        0.267  22         0
## 34                        0.188  28         0
## 35                        0.512  45         0
## 36                        0.966  33         0
## 37                        0.420  35         0
## 38                        0.665  46         1
## 39                        0.503  27         1
## 40                        1.390  56         1
## 41                        0.271  26         0
## 42                        0.696  37         0
## 43                        0.235  48         0
## 44                        0.721  54         1
## 45                        0.294  40         0
## 46                        1.893  25         1
## 47                        0.564  29         0
## 48                        0.586  22         0
## 49                        0.344  31         1
## 50                        0.305  24         0
## 51                        0.491  22         0
## 52                        0.526  26         0
## 53                        0.342  30         0
## 54                        0.467  58         1
## 55                        0.718  42         0
## 56                        0.248  21         0
## 57                        0.254  41         1
## 58                        0.962  31         0
## 59                        1.781  44         0
## 60                        0.173  22         0
## 61                        0.304  21         0
## 62                        0.270  39         1
## 63                        0.587  36         0
## 64                        0.699  24         0
## 65                        0.258  42         1
## 66                        0.203  32         0
## 67                        0.855  38         1
## 68                        0.845  54         0
## 69                        0.334  25         0
## 70                        0.189  27         0
## 71                        0.867  28         1
## 72                        0.411  26         0
## 73                        0.583  42         1
## 74                        0.231  23         0
## 75                        0.396  22         0
## 76                        0.140  22         0
```

```
## 77                        0.391  41        0
## 78                        0.370  27        0
## 79                        0.270  26        1
## 80                        0.307  24        0
## 81                        0.140  22        0
## 82                        0.102  22        0
## 83                        0.767  36        0
## 84                        0.237  22        0
## 85                        0.227  37        1
## 86                        0.698  27        0
## 87                        0.178  45        0
## 88                        0.324  26        0
## 89                        0.153  43        1
## 90                        0.165  24        0
## 91                        0.258  21        0
## 92                        0.443  34        0
## 93                        0.261  42        0
## 94                        0.277  60        1
## 95                        0.761  21        0
## 96                        0.255  40        0
## 97                        0.130  24        0
## 98                        0.323  22        0
## 99                        0.356  23        0
## 100                       0.325  31        1
## 101                       1.222  33        1
## 102                       0.179  22        0
## 103                       0.262  21        0
## 104                       0.283  24        0
## 105                       0.930  27        0
## 106                       0.801  21        0
## 107                       0.207  27        0
## 108                       0.287  37        0
## 109                       0.336  25        0
## 110                       0.247  24        1
## 111                       0.199  24        1
## 112                       0.543  46        1
## 113                       0.192  23        0
## 114                       0.391  25        0
## 115                       0.588  39        1
## 116                       0.539  61        1
## 117                       0.220  38        1
## 118                       0.654  25        0
## 119                       0.443  22        0
## 120                       0.223  21        0
## 121                       0.759  25        1
## 122                       0.260  24        0
## 123                       0.404  23        0
## 124                       0.186  69        0
## 125                       0.278  23        1
## 126                       0.496  26        1
## 127                       0.452  30        0
## 128                       0.261  23        0
## 129                       0.403  40        1
## 130                       0.741  62        1
```

```
## 131                          0.361  33        1
## 132                          1.114  33        1
## 133                          0.356  30        1
## 134                          0.457  39        0
## 135                          0.647  26        0
## 136                          0.088  31        0
## 137                          0.597  21        0
## 138                          0.532  22        0
## 139                          0.703  29        0
## 140                          0.159  28        0
## 141                          0.268  55        0
## 142                          0.286  38        0
## 143                          0.318  22        0
## 144                          0.272  42        1
## 145                          0.237  23        0
## 146                          0.572  21        0
## 147                          0.096  41        0
## 148                          1.400  34        0
## 149                          0.218  65        0
## 150                          0.085  22        0
## 151                          0.399  24        0
## 152                          0.432  37        0
## 153                          1.189  42        1
## 154                          0.687  23        0
## 155                          0.137  43        1
## 156                          0.337  36        1
## 157                          0.637  21        0
## 158                          0.833  23        0
## 159                          0.229  22        0
## 160                          0.817  47        1
## 161                          0.294  36        0
## 162                          0.204  45        0
## 163                          0.167  27        0
## 164                          0.368  21        0
## 165                          0.743  32        1
## 166                          0.722  41        1
## 167                          0.256  22        0
## 168                          0.709  34        0
## 169                          0.471  29        0
## 170                          0.495  29        0
## 171                          0.180  36        1
## 172                          0.542  29        1
## 173                          0.773  25        0
## 174                          0.678  23        0
## 175                          0.370  33        0
## 176                          0.719  36        1
## 177                          0.382  42        0
## 178                          0.319  26        1
## 179                          0.190  47        0
## 180                          0.956  37        1
## 181                          0.084  32        0
## 182                          0.725  23        0
## 183                          0.299  21        0
## 184                          0.268  27        0
```

```
## 185                          0.244  40        0
## 186                          0.745  41        1
## 187                          0.615  60        1
## 188                          1.321  33        1
## 189                          0.640  31        1
## 190                          0.361  25        1
## 191                          0.142  21        0
## 192                          0.374  40        0
## 193                          0.383  36        1
## 194                          0.578  40        1
## 195                          0.136  42        0
## 196                          0.395  29        1
## 197                          0.187  21        0
## 198                          0.678  23        1
## 199                          0.905  26        1
## 200                          0.150  29        1
## 201                          0.874  21        0
## 202                          0.236  28        0
## 203                          0.787  32        0
## 204                          0.235  27        0
## 205                          0.324  55        0
## 206                          0.407  27        0
## 207                          0.605  57        1
## 208                          0.151  52        1
## 209                          0.289  21        0
## 210                          0.355  41        1
## 211                          0.290  25        0
## 212                          0.375  24        0
## 213                          0.164  60        0
## 214                          0.431  24        1
## 215                          0.260  36        1
## 216                          0.742  38        1
## 217                          0.514  25        1
## 218                          0.464  32        0
## 219                          1.224  32        1
## 220                          0.261  41        1
## 221                          1.072  21        1
## 222                          0.805  66        1
## 223                          0.209  37        0
## 224                          0.687  61        0
## 225                          0.666  26        0
## 226                          0.101  22        0
## 227                          0.198  26        0
## 228                          0.652  24        1
## 229                          2.329  31        0
## 230                          0.089  24        0
## 231                          0.645  22        1
## 232                          0.238  46        1
## 233                          0.583  22        0
## 234                          0.394  29        0
## 235                          0.293  23        0
## 236                          0.479  26        1
## 237                          0.586  51        1
## 238                          0.686  23        1
```

```
## 239                         0.831  32        1
## 240                         0.582  27        0
## 241                         0.192  21        0
## 242                         0.446  22        0
## 243                         0.402  22        1
## 244                         1.318  33        1
## 245                         0.329  29        0
## 246                         1.213  49        1
## 247                         0.258  41        0
## 248                         0.427  23        0
## 249                         0.282  34        0
## 250                         0.143  23        0
## 251                         0.380  42        0
## 252                         0.284  27        0
## 253                         0.249  24        0
## 254                         0.238  25        0
## 255                         0.926  44        1
## 256                         0.543  21        1
## 257                         0.557  30        0
## 258                         0.092  25        0
## 259                         0.655  24        0
## 260                         1.353  51        1
## 261                         0.299  34        0
## 262                         0.761  27        1
## 263                         0.612  24        0
## 264                         0.200  63        0
## 265                         0.226  35        1
## 266                         0.997  43        0
## 267                         0.933  25        1
## 268                         1.101  24        0
## 269                         0.078  21        0
## 270                         0.240  28        1
## 271                         1.136  38        1
## 272                         0.128  21        0
## 273                         0.254  40        0
## 274                         0.422  21        0
## 275                         0.251  52        0
## 276                         0.677  25        0
## 277                         0.296  29        1
## 278                         0.454  23        0
## 279                         0.744  57        0
## 280                         0.881  22        0
## 281                         0.334  28        1
## 282                         0.280  39        0
## 283                         0.262  37        0
## 284                         0.165  47        1
## 285                         0.259  52        1
## 286                         0.647  51        0
## 287                         0.619  34        0
## 288                         0.808  29        1
## 289                         0.340  26        0
## 290                         0.263  33        0
## 291                         0.434  21        0
## 292                         0.757  25        1
```

```
## 293                         1.224  31        1
## 294                         0.613  24        1
## 295                         0.254  65        0
## 296                         0.692  28        0
## 297                         0.337  29        1
## 298                         0.520  24        0
## 299                         0.412  46        1
## 300                         0.840  58        0
## 301                         0.839  30        1
## 302                         0.422  25        1
## 303                         0.156  35        0
## 304                         0.209  28        1
## 305                         0.207  37        0
## 306                         0.215  29        0
## 307                         0.326  47        1
## 308                         0.143  21        0
## 309                         1.391  25        1
## 310                         0.875  30        1
## 311                         0.313  41        0
## 312                         0.605  22        0
## 313                         0.433  27        1
## 314                         0.626  25        0
## 315                         1.127  43        1
## 316                         0.315  26        0
## 317                         0.284  30        0
## 318                         0.345  29        1
## 319                         0.150  28        0
## 320                         0.129  59        1
## 321                         0.527  31        0
## 322                         0.197  25        1
## 323                         0.254  36        1
## 324                         0.731  43        1
## 325                         0.148  21        0
## 326                         0.123  24        0
## 327                         0.692  30        1
## 328                         0.200  37        0
## 329                         0.127  23        1
## 330                         0.122  37        0
## 331                         1.476  46        0
## 332                         0.166  25        0
## 333                         0.282  41        1
## 334                         0.137  44        0
## 335                         0.260  22        0
## 336                         0.259  26        0
## 337                         0.932  44        0
## 338                         0.343  44        1
## 339                         0.893  33        1
## 340                         0.331  41        1
## 341                         0.472  22        0
## 342                         0.673  36        0
## 343                         0.389  22        0
## 344                         0.290  33        0
## 345                         0.485  57        0
## 346                         0.349  49        0
```

```
## 347                     0.654  22         0
## 348                     0.187  23         0
## 349                     0.279  26         0
## 350                     0.346  37         1
## 351                     0.237  29         0
## 352                     0.252  30         0
## 353                     0.243  46         0
## 354                     0.580  24         0
## 355                     0.559  21         0
## 356                     0.302  49         1
## 357                     0.962  28         1
## 358                     0.569  44         1
## 359                     0.378  48         0
## 360                     0.875  29         1
## 361                     0.583  29         1
## 362                     0.207  63         0
## 363                     0.305  65         0
## 364                     0.520  67         1
## 365                     0.385  30         0
## 366                     0.499  30         0
## 367                     0.368  29         1
## 368                     0.252  21         0
## 369                     0.306  22         0
## 370                     0.234  45         1
## 371                     2.137  25         1
## 372                     1.731  21         0
## 373                     0.545  21         0
## 374                     0.225  25         0
## 375                     0.816  28         0
## 376                     0.528  58         1
## 377                     0.299  22         0
## 378                     0.509  22         0
## 379                     0.238  32         1
## 380                     1.021  35         0
## 381                     0.821  24         0
## 382                     0.236  22         0
## 383                     0.947  21         0
## 384                     1.268  25         0
## 385                     0.221  25         0
## 386                     0.205  24         0
## 387                     0.660  35         1
## 388                     0.239  45         1
## 389                     0.452  58         1
## 390                     0.949  28         0
## 391                     0.444  42         0
## 392                     0.340  27         1
## 393                     0.389  21         0
## 394                     0.463  37         0
## 395                     0.803  31         1
## 396                     1.600  25         0
## 397                     0.944  39         0
## 398                     0.196  22         1
## 399                     0.389  25         0
## 400                     0.241  25         1
```

```
## 401                        0.161  31       1
## 402                        0.151  55       0
## 403                        0.286  35       1
## 404                        0.280  38       0
## 405                        0.135  41       1
## 406                        0.520  26       0
## 407                        0.376  46       1
## 408                        0.336  25       0
## 409                        1.191  39       1
## 410                        0.702  28       1
## 411                        0.674  28       0
## 412                        0.528  25       0
## 413                        1.076  22       0
## 414                        0.256  21       0
## 415                        0.534  21       1
## 416                        0.258  22       1
## 417                        1.095  22       0
## 418                        0.554  37       1
## 419                        0.624  27       0
## 420                        0.219  28       1
## 421                        0.507  26       0
## 422                        0.561  21       0
## 423                        0.496  21       0
## 424                        0.421  21       0
## 425                        0.516  36       1
## 426                        0.264  31       1
## 427                        0.256  25       0
## 428                        0.328  38       1
## 429                        0.284  26       0
## 430                        0.233  43       1
## 431                        0.108  23       0
## 432                        0.551  38       0
## 433                        0.527  22       0
## 434                        0.167  29       0
## 435                        1.138  36       0
## 436                        0.205  29       1
## 437                        0.244  41       0
## 438                        0.434  28       0
## 439                        0.147  21       0
## 440                        0.727  31       0
## 441                        0.435  41       1
## 442                        0.497  22       0
## 443                        0.230  24       0
## 444                        0.955  33       1
## 445                        0.380  30       1
## 446                        2.420  25       1
## 447                        0.658  28       0
## 448                        0.330  26       0
## 449                        0.510  22       1
## 450                        0.285  26       0
## 451                        0.415  23       0
## 452                        0.542  23       1
## 453                        0.381  25       0
## 454                        0.832  72       0
```

```
## 455                         0.498  24         0
## 456                         0.212  38         1
## 457                         0.687  62         0
## 458                         0.364  24         0
## 459                         1.001  51         1
## 460                         0.460  81         0
## 461                         0.733  48         0
## 462                         0.416  26         0
## 463                         0.705  39         0
## 464                         0.258  37         0
## 465                         1.022  34         0
## 466                         0.452  21         0
## 467                         0.269  22         0
## 468                         0.600  25         0
## 469                         0.183  38         1
## 470                         0.571  27         0
## 471                         0.607  28         0
## 472                         0.170  22         0
## 473                         0.259  22         0
## 474                         0.210  50         0
## 475                         0.126  24         0
## 476                         0.231  59         0
## 477                         0.711  29         1
## 478                         0.466  31         0
## 479                         0.162  39         0
## 480                         0.419  63         0
## 481                         0.344  35         1
## 482                         0.197  29         0
## 483                         0.306  28         0
## 484                         0.233  23         0
## 485                         0.630  31         1
## 486                         0.365  24         1
## 487                         0.536  21         0
## 488                         1.159  58         0
## 489                         0.294  28         0
## 490                         0.551  67         0
## 491                         0.629  24         0
## 492                         0.292  42         0
## 493                         0.145  33         0
## 494                         1.144  45         1
## 495                         0.174  22         0
## 496                         0.304  66         0
## 497                         0.292  30         0
## 498                         0.547  25         0
## 499                         0.163  55         1
## 500                         0.839  39         0
## 501                         0.313  21         0
## 502                         0.267  28         0
## 503                         0.727  41         1
## 504                         0.738  41         0
## 505                         0.238  40         0
## 506                         0.263  38         0
## 507                         0.314  35         1
## 508                         0.692  21         0
```

```
## 509                         0.968  21        0
## 510                         0.409  64        0
## 511                         0.297  46        1
## 512                         0.207  21        0
## 513                         0.200  58        0
## 514                         0.525  22        0
## 515                         0.154  24        0
## 516                         0.268  28        1
## 517                         0.771  53        1
## 518                         0.304  51        0
## 519                         0.180  41        0
## 520                         0.582  60        0
## 521                         0.187  25        0
## 522                         0.305  26        0
## 523                         0.189  26        0
## 524                         0.652  45        1
## 525                         0.151  24        0
## 526                         0.444  21        0
## 527                         0.299  21        0
## 528                         0.107  24        0
## 529                         0.493  22        0
## 530                         0.660  31        0
## 531                         0.717  22        0
## 532                         0.686  24        0
## 533                         0.917  29        0
## 534                         0.501  31        0
## 535                         1.251  24        0
## 536                         0.302  23        1
## 537                         0.197  46        0
## 538                         0.735  67        0
## 539                         0.804  23        0
## 540                         0.968  32        1
## 541                         0.661  43        1
## 542                         0.549  27        1
## 543                         0.825  56        1
## 544                         0.159  25        0
## 545                         0.365  29        0
## 546                         0.423  37        1
## 547                         1.034  53        1
## 548                         0.160  28        0
## 549                         0.341  50        0
## 550                         0.680  37        0
## 551                         0.204  21        0
## 552                         0.591  25        0
## 553                         0.247  66        0
## 554                         0.422  23        0
## 555                         0.471  28        0
## 556                         0.161  37        0
## 557                         0.218  30        0
## 558                         0.237  58        0
## 559                         0.126  42        0
## 560                         0.300  35        0
## 561                         0.121  54        1
## 562                         0.502  28        1
```

```
## 563                          0.401  24         0
## 564                          0.497  32         0
## 565                          0.601  27         0
## 566                          0.748  22         0
## 567                          0.412  21         0
## 568                          0.085  46         0
## 569                          0.338  37         0
## 570                          0.203  33         1
## 571                          0.270  39         0
## 572                          0.268  21         0
## 573                          0.430  22         0
## 574                          0.198  22         0
## 575                          0.892  23         0
## 576                          0.280  25         0
## 577                          0.813  35         0
## 578                          0.693  21         1
## 579                          0.245  36         0
## 580                          0.575  62         1
## 581                          0.371  21         1
## 582                          0.206  27         0
## 583                          0.259  62         0
## 584                          0.190  42         0
## 585                          0.687  52         1
## 586                          0.417  22         0
## 587                          0.129  41         1
## 588                          0.249  29         0
## 589                          1.154  52         1
## 590                          0.342  25         0
## 591                          0.925  45         1
## 592                          0.175  24         0
## 593                          0.402  44         1
## 594                          1.699  25         0
## 595                          0.733  34         0
## 596                          0.682  22         1
## 597                          0.194  46         0
## 598                          0.559  21         0
## 599                          0.088  38         1
## 600                          0.407  26         0
## 601                          0.400  24         0
## 602                          0.190  28         0
## 603                          0.100  30         0
## 604                          0.692  54         1
## 605                          0.212  36         1
## 606                          0.514  21         0
## 607                          1.258  22         1
## 608                          0.482  25         0
## 609                          0.270  27         0
## 610                          0.138  23         0
## 611                          0.292  24         0
## 612                          0.593  36         1
## 613                          0.787  40         1
## 614                          0.878  26         0
## 615                          0.557  50         1
## 616                          0.207  27         0
```

```
## 617                          0.157  30        0
## 618                          0.257  23        0
## 619                          1.282  50        1
## 620                          0.141  24        1
## 621                          0.246  28        0
## 622                          1.698  28        0
## 623                          1.461  45        0
## 624                          0.347  21        0
## 625                          0.158  21        0
## 626                          0.362  29        0
## 627                          0.206  21        0
## 628                          0.393  21        0
## 629                          0.144  45        0
## 630                          0.148  21        0
## 631                          0.732  34        1
## 632                          0.238  24        0
## 633                          0.343  23        0
## 634                          0.115  22        0
## 635                          0.167  31        0
## 636                          0.465  38        1
## 637                          0.153  48        0
## 638                          0.649  23        0
## 639                          0.871  32        1
## 640                          0.149  28        0
## 641                          0.695  27        0
## 642                          0.303  24        0
## 643                          0.178  50        1
## 644                          0.610  31        0
## 645                          0.730  27        0
## 646                          0.134  30        0
## 647                          0.447  33        1
## 648                          0.455  22        1
## 649                          0.260  42        1
## 650                          0.133  23        0
## 651                          0.234  23        0
## 652                          0.466  27        0
## 653                          0.269  28        0
## 654                          0.455  27        0
## 655                          0.142  22        0
## 656                          0.240  25        1
## 657                          0.155  22        0
## 658                          1.162  41        0
## 659                          0.190  51        0
## 660                          1.292  27        1
## 661                          0.182  54        0
## 662                          1.394  22        1
## 663                          0.165  43        1
## 664                          0.637  40        1
## 665                          0.245  40        1
## 666                          0.217  24        0
## 667                          0.235  70        1
## 668                          0.141  40        1
## 669                          0.430  43        0
## 670                          0.164  45        0
```

```
## 671                          0.631  49        0
## 672                          0.551  21        0
## 673                          0.285  47        0
## 674                          0.880  22        0
## 675                          0.587  68        0
## 676                          0.328  31        1
## 677                          0.230  53        1
## 678                          0.263  25        0
## 679                          0.127  25        1
## 680                          0.614  23        0
## 681                          0.332  22        0
## 682                          0.364  26        1
## 683                          0.366  22        0
## 684                          0.536  27        1
## 685                          0.640  69        0
## 686                          0.591  25        0
## 687                          0.314  22        0
## 688                          0.181  29        0
## 689                          0.828  23        0
## 690                          0.335  46        1
## 691                          0.856  34        0
## 692                          0.257  44        1
## 693                          0.886  23        0
## 694                          0.439  43        1
## 695                          0.191  25        0
## 696                          0.128  43        1
## 697                          0.268  31        1
## 698                          0.253  22        0
## 699                          0.598  28        0
## 700                          0.904  26        0
## 701                          0.483  26        0
## 702                          0.565  49        1
## 703                          0.905  52        1
## 704                          0.304  41        0
## 705                          0.118  27        0
## 706                          0.177  28        0
## 707                          0.261  30        1
## 708                          0.176  22        0
## 709                          0.148  45        1
## 710                          0.674  23        1
## 711                          0.295  24        0
## 712                          0.439  40        0
## 713                          0.441  38        1
## 714                          0.352  21        0
## 715                          0.121  32        0
## 716                          0.826  34        1
## 717                          0.970  31        1
## 718                          0.595  56        0
## 719                          0.415  24        0
## 720                          0.378  52        1
## 721                          0.317  34        0
## 722                          0.289  21        0
## 723                          0.349  42        1
## 724                          0.251  42        0
```

```
## 725                          0.265  45        0
## 726                          0.236  38        0
## 727                          0.496  25        0
## 728                          0.433  22        0
## 729                          0.326  22        0
## 730                          0.141  22        0
## 731                          0.323  34        1
## 732                          0.259  22        1
## 733                          0.646  24        1
## 734                          0.426  22        0
## 735                          0.560  53        0
## 736                          0.284  28        0
## 737                          0.515  21        0
## 738                          0.600  42        0
## 739                          0.453  21        0
## 740                          0.293  42        1
## 741                          0.785  48        1
## 742                          0.400  26        0
## 743                          0.219  22        0
## 744                          0.734  45        1
## 745                          1.174  39        0
## 746                          0.488  46        0
## 747                          0.358  27        1
## 748                          1.096  32        0
## 749                          0.408  36        1
## 750                          0.178  50        1
## 751                          1.182  22        1
## 752                          0.261  28        0
## 753                          0.223  25        0
## 754                          0.222  26        1
## 755                          0.443  45        1
## 756                          1.057  37        1
## 757                          0.391  39        0
## 758                          0.258  52        1
## 759                          0.197  26        0
## 760                          0.278  66        1
## 761                          0.766  22        0
## 762                          0.403  43        1
## 763                          0.142  33        0
## 764                          0.171  63        0
## 765                          0.340  27        0
## 766                          0.245  30        0
## 767                          0.349  47        1
## 768                          0.315  23        0
```

First, we categorized the BMI category into Underweight, Normal, Overweight, and Obesity.

** Obesity vs Outcome**

```
counts <- as.data.frame(with(df.temp, table(BMI, Outcome)))
colnames(counts) <- c("BMI", "Outcome", "Freq")

counts <- counts %>%
  group_by(BMI) %>%
```

```r
  mutate(prop = Freq / sum(Freq))

outcome_colors <- c("1" = 'salmon', "0" = 'lightyellow')

plot_ly(
  data = counts,
  x = ~BMI,
  y = ~prop,
  color = ~Outcome,
  colors = outcome_colors,
  type = "bar"
) %>%
  layout(
    title = "Proportion of Outcomes by BMI Category",
    barmode = "stack",
    xaxis = list(title = "BMI", categoryorder = "array", categoryarray = c("Underweight", "Normal", "Ove
    yaxis = list(title = "Frequency"),
    legend = list(title = list(text = "Outcome", font = list(size = 15, color = "black")))
  )
```

The graph show that there is no Underweight individual that has diabetes, only 6.9% individuals with Normal BMI that have diabetes, 22.3% Overweight individuals in fact have diabetes, and lastly, 45.8% individuals that have obesity also suffer from diabetes.

Here we can conclude that individuals that classified as Overweight and Obese have significantly higher risk of having diabetes compared to those with a Normal BMI or who are Underweight. So it is crucial to maintain a healthy BMI to reduce the risk of diabetes.

Age vs Outcome

```r
age_counts <- as.data.frame(with(df, table(Age, Outcome)))
colnames(age_counts) <- c("Age", "Outcome", "Freq")

age_counts <- age_counts %>%
  group_by(Age) %>%
  mutate(prop = Freq / sum(Freq))

outcome_colors <- c('0' = '#FFFDD0', '1' = 'lightblue')

# Plot the stacked bar chart with Plotly
plot_ly(
  data = age_counts,
  x = ~Age,
  y = ~prop,
  color = ~Outcome,
  colors = outcome_colors,
  type = "bar"
) %>%
  layout(
    barmode = "stack",
    xaxis = list(title = "Age"),
    yaxis = list(title = "Frequency")
  )
```

There is higher proportion of outcome 0 (non-diabetical) in younger individuals between ages 20 to 35, with gradual decrease in frequencies in between. Between the age of 36 to 54, the outcome 1 (diabetical) start to become dominant where it is more common than outcome 0. From age 55 and above the green bars (diabetical) is continue to dominate.

This suggest that non-diabetical is more relevant to younger people, while diabetes issue becomes increasingly common as people starting to age, dominating in the older age groups.

How heredity affect diabetes?

DiabetesPedigreeFunction vs Outcome

```
pedigree_counts <- as.data.frame(with(df, table(DiabetesPedigreeFunction, Outcome)))
colnames(pedigree_counts) <- c("DiabetesPedigreeFunction", "Outcome", "Freq")

pedigree_counts <- pedigree_counts %>%
  group_by(DiabetesPedigreeFunction) %>%
  mutate(propp = Freq / sum(Freq))

# Plotting the scatter plot
plot_ly(
  data = pedigree_counts,
  x = ~DiabetesPedigreeFunction,
  y = ~propp,
  type = "bar",
  #mode = "lines+markers",
  jitter = 0.1,
  color = ~Outcome,
  colors = c('0' = 'lightyellow', '1' = 'red')
) %>%
layout(
  title = "Proportion of Outcomes by Diabetes Pedigree Function",
  xaxis = list(title = "Diabetes Pedigree Function"),
  yaxis = list(title = "Frequency")
)
```

From the graph we can see, at lower values of DiabetesPedigreeFunction, there are more yellow bars, showing there is a higher proportion of non-diabetical individuals. But as the DiabetesPedigreeFunction value increase the proportion of diabetic individuals (red bars) is also increasing.

We can conclude that DiabetesPedigreeFunction is in fact affect the Outcome of diabetes, with a positive linear relationship.

Why DiabetesS is a problem?

```
count_data <- df %>%
  group_by(BloodPressure, Outcome) %>%
  summarise(Count = n()) %>%
  ungroup()

count_data_0 <- count_data %>% filter(Outcome == 0)
count_data_1 <- count_data %>% filter(Outcome == 1)
```

```r
fig <- plot_ly()

fig <- fig %>% add_trace(
  x = ~count_data_0$BloodPressure,
  y = ~count_data_0$Count,
  type = "scatter",
  mode = "lines",
  name = "No Diabetes",
  line = list(color = 'darkred') # Darker shade of red
)

fig <- fig %>% add_trace(
  x = ~count_data_1$BloodPressure,
  y = ~count_data_1$Count,
  type = "scatter",
  mode = "lines+markers",
  name = "Has Diabetes",
  line = list(color = 'red')
)

fig <- fig %>% layout(
  title = "Distributions Blood Pressure by Outcome",
  xaxis = list(title = "Blood Pressure"),
  yaxis = list(title = "Count"),
  legend = list(title = list(text = 'Outcome'))
)

fig
```

From the graph, we can see that both groups have a quite similar range of blood pressure values, with non-diabetical's bloodPressure range from 24 to 122 mm hg, and diabetical's bloodPressure range from 30 to 144 mm hg. This shows that there is possible increase of blood pressure in individuals with diabetes and it is crucial for people with diabetes to continue monitoring their blood pressure to prevent further complications.

d) DATA ANOMALIES/OUTLIERS

ThreeSigma, Hampel, Boxplot Rule

```r
# nMiss for missing value, nOut for outlier
# lowLim upLim, lower and upper outlier detection limits

find_outliers_summary <- function(x){
  outliers <- FindOutliers(x)
  return(outliers$summary)
}

summary_list <- lapply(df, find_outliers_summary)
summary_list
```

```
## $Pregnancies
```

```
##        method   n nMiss nOut    lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0    4 -6.263682 13.95379      0     13
## 2      Hampel 768     0   23 -5.895600 11.89560      0     11
## 3 BoxplotRule 768     0    4 -1.500000 13.50000      0     13
##
## $Glucose
##        method   n nMiss nOut   lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0    0 30.37356 212.9897     44    199
## 2      Hampel 768     0    0 28.04400 205.9560     44    199
## 3 BoxplotRule 768     0   36 79.50000 201.0000     80    199
##
## $BloodPressure
##        method   n nMiss nOut   lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0    8 35.90701 108.6026     38    108
## 2      Hampel 768     0   10 36.41760 107.5824     38    106
## 3 BoxplotRule 768     0   63 56.00000 104.0000     56    104
##
## $SkinThickness
##        method   n nMiss nOut     lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0    4 -0.3524075 55.02168      7     54
## 2      Hampel 768     0   31  0.7610000 45.23900      7     45
## 3 BoxplotRule 768     0  124 18.5000000 45.50000     19     45
##
## $Insulin
##        method   n nMiss nOut     lowLim     upLim minNom maxNom
## 1  ThreeSigma 768     0   19 -221.99045 411.29514     14    402
## 2      Hampel 768     0  316   -5.44435  67.94435     14     67
## 3 BoxplotRule 768     0   49  -17.87500 272.37500     14    272
##
## $BMI
##        method   n nMiss nOut   lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0    5 11.82468 53.07693   18.2   52.9
## 2      Hampel 768     0    8 11.98490 52.01510   18.2   50.0
## 3 BoxplotRule 768     0   58 22.95000 50.25000   23.0   50.0
##
## $DiabetesPedigreeFunction
##        method   n nMiss nOut     lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0   11 -0.5221095 1.465862  0.078  1.461
## 2      Hampel 768     0   40 -0.3725065 1.117506  0.078  1.114
## 3 BoxplotRule 768     0   29  0.0525000 1.200000  0.078  1.191
##
## $Age
##        method   n nMiss nOut    lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0    5 -2.039809 68.52158     21     68
## 2      Hampel 768     0   27 -2.134600 60.13460     21     60
## 3 BoxplotRule 768     0    9 15.500000 66.50000     21     66
##
## $Outcome
##        method   n nMiss nOut    lowLim    upLim minNom maxNom
## 1  ThreeSigma 768     0    0 -1.081896 1.779812      0      1
## 2      Hampel 768     0  268  0.000000 0.000000      0      0
## 3 BoxplotRule 768     0    0 -0.500000 2.500000      0      1
```

Number of outliers using ThreeSigma method:

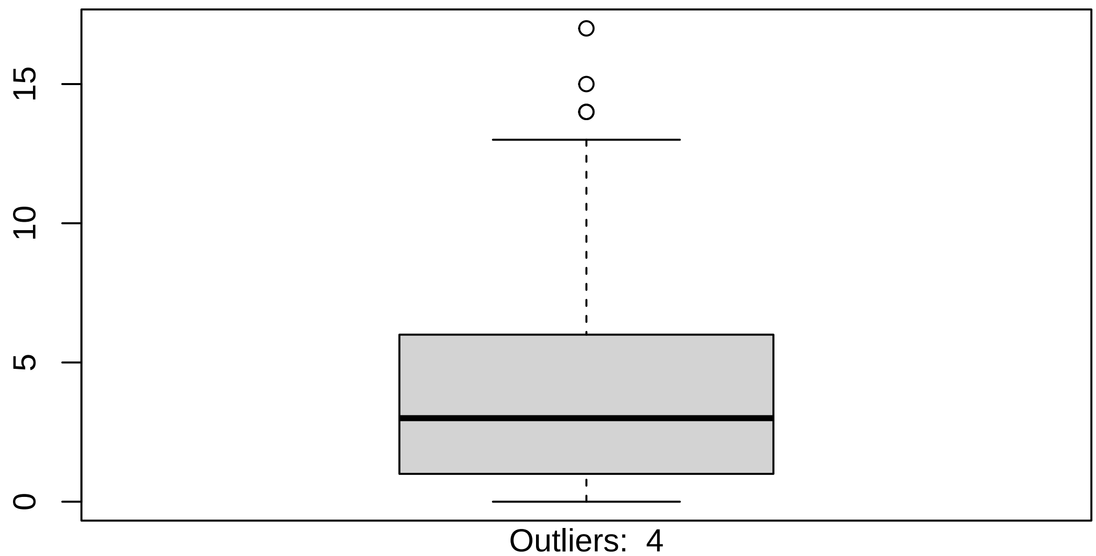
- Pregnancies : 4
- Glucose : 0
- BloodPressure : 8
- SkinThickness : 4
- Insulin : 19
- BMI : 5
- DiabetesPedigreeFunction : 11
- Age : 5 - Outcome : 0

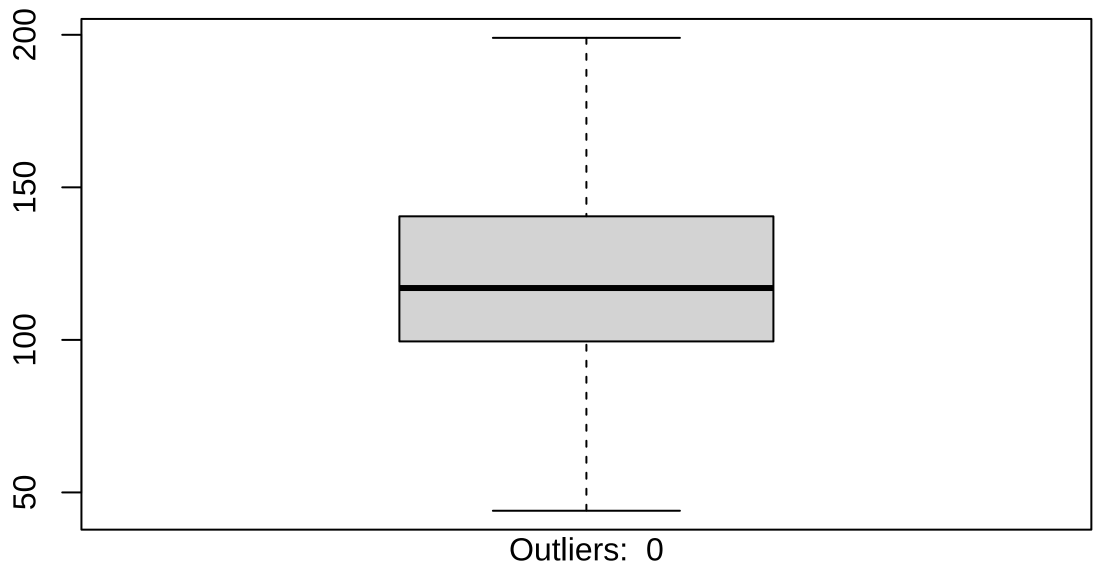
Number of outliers using Hampel method:
- Pregnancies : 23
- Glucose : 0
- BloodPressure : 10
- SkinThickness : 31
- Insulin : 316
- BMI : 8
- DiabetesPedigreeFunction : 40
- Age : 27
- Outcome : 268

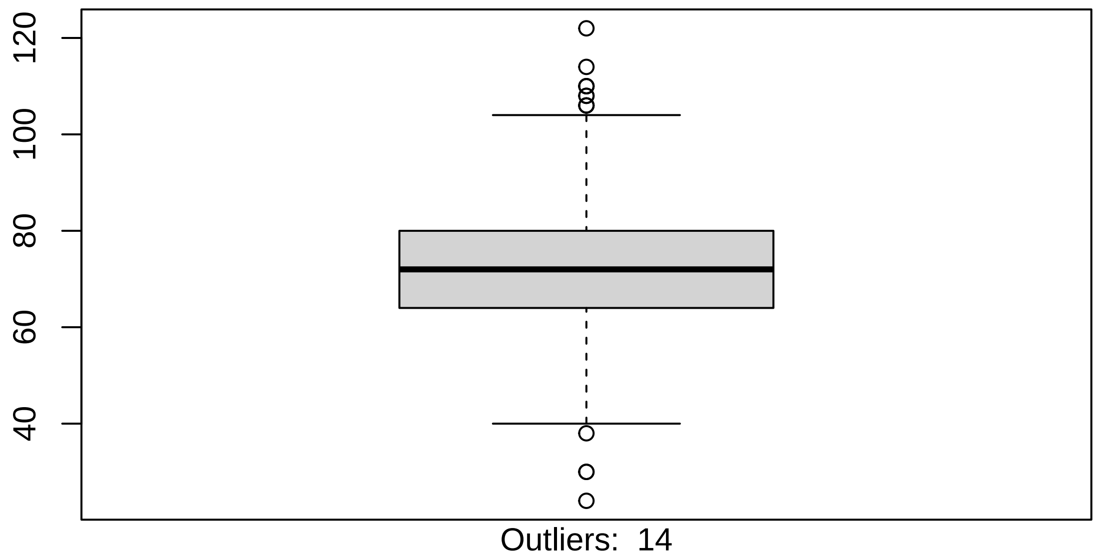
Number of outliers using BoxplotRule method:
- Pregnancies : 4
- Glucose : 36
- BloodPressure : 63
- SkinThickness : 124
- Insulin : 49
- BMI : 58
- DiabetesPedigreeFunction : 29
- Age : 9 - Outcome : 0

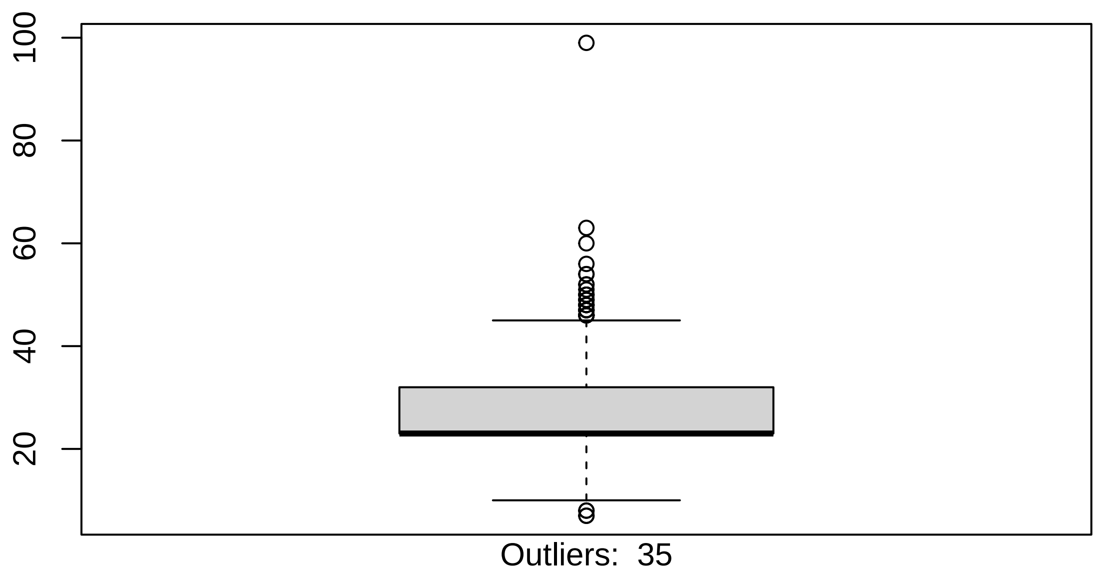
using boxplot

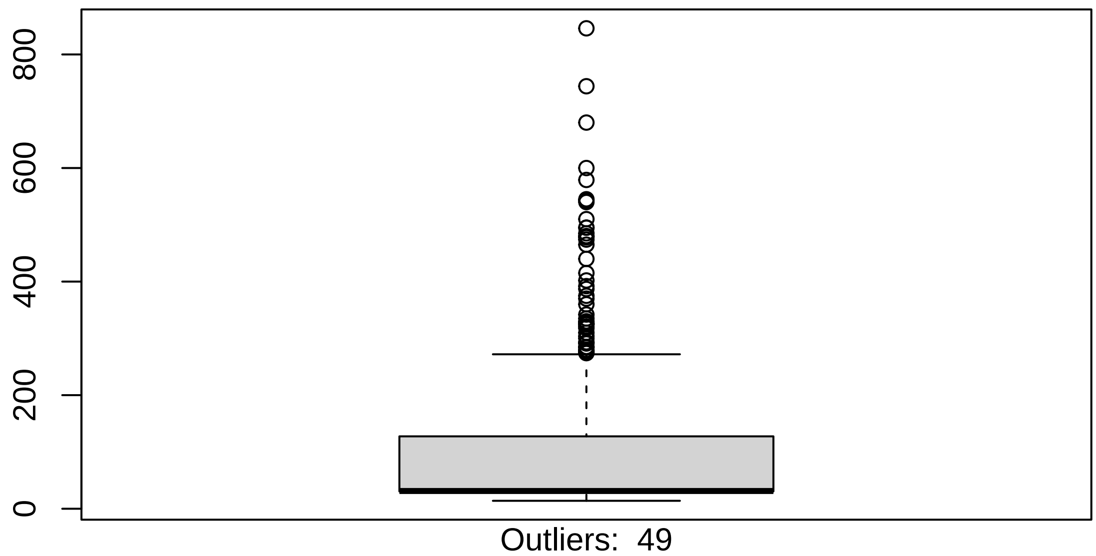
```r
count_outliers <- function(x) {
  bp <- boxplot(x, plot = FALSE)
  out <- length(bp$out)
  boxplot(x)
  mtext(paste("Outliers: ", out), side = 1)
}

sapply(df, count_outliers)
```

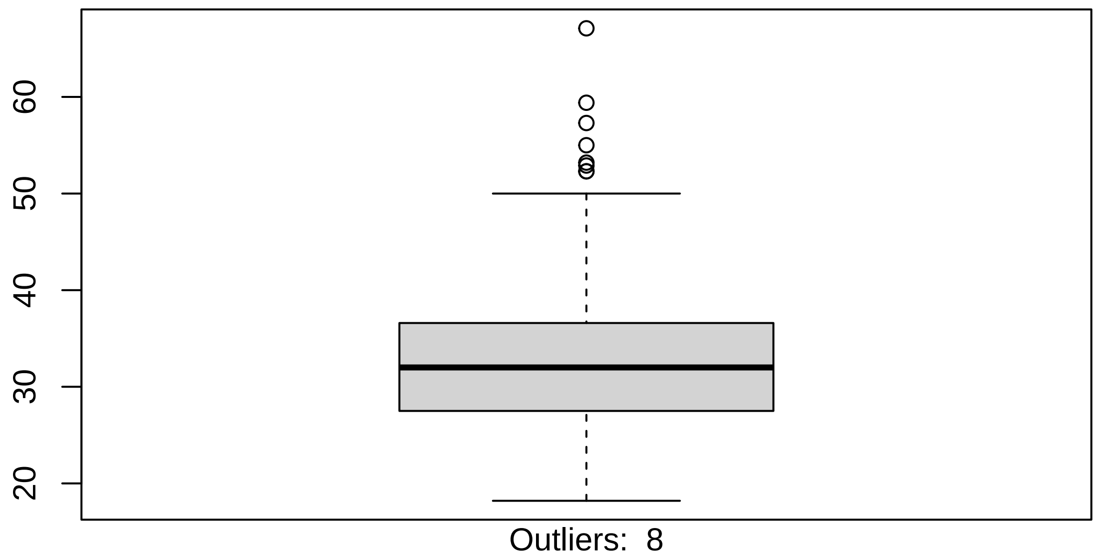

15
10
5
0
Outliers: 4

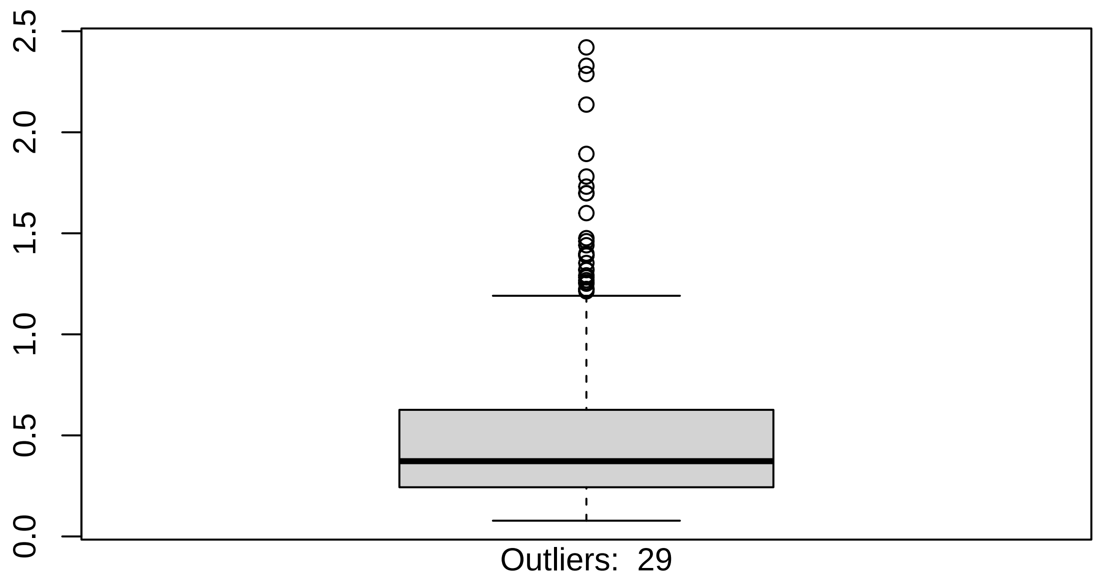

200
150
100
50
Outliers: 0

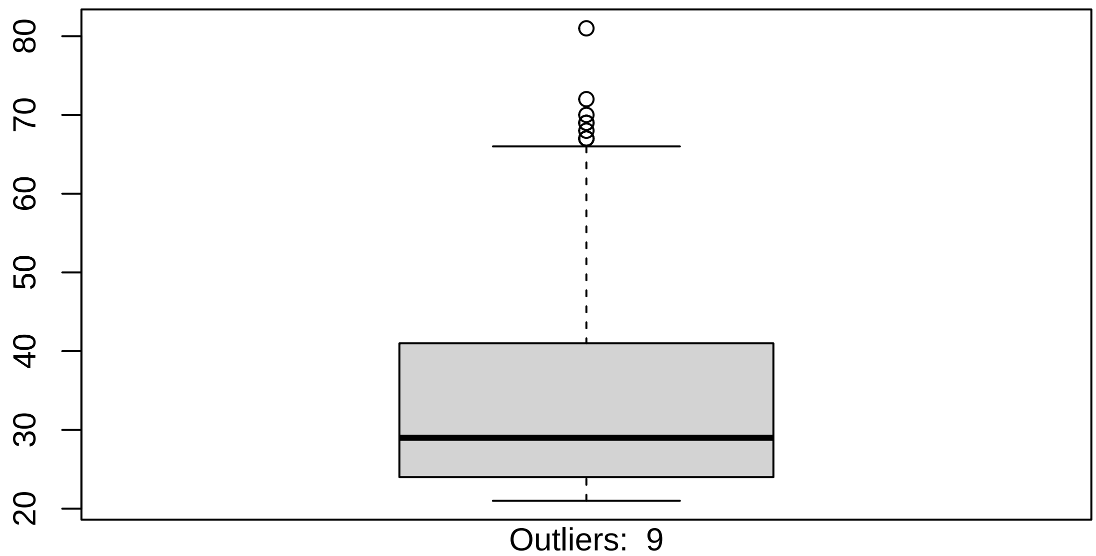

120
100
80
60
40
Outliers: 14


100
80
60
40
20
Outliers: 35


800
600
400
200
0
Outliers: 49


20
30
40
50
60
Outliers: 8


2.5
2.0
1.5
1.0
0.5
0.0
Outliers: 29



20
30
40
50
60
70
80
Outliers: 9


```
## $Pregnancies
## NULL
##
## $Glucose
## NULL
##
## $BloodPressure
## NULL
##
## $SkinThickness
## NULL
##
## $Insulin
## NULL
##
## $BMI
## NULL
##
## $DiabetesPedigreeFunction
## NULL
##
## $Age
## NULL
##
## $Outcome
## NULL
```

The numbers of outlier using boxplot:
- Pregnancies : 4
- Glucose : 0
- BloodPressure : 14
- SkinThickness : 35
- Insulin : 49
- BMI : 8
- DiabetesPedigreeFunction : 29
- Age : 9 - Outcome : 0

Using formula

```
count_outliers2 <- function(x) {
  q1 <- quantile(x, p = 0.25)
  q3 <- quantile(x, p = 0.75)
  iqr <- IQR(x)
  # determine outliers
  outliers <- ifelse(x < q1 - 1.5 * iqr | x > q3 + 1.5 * iqr, TRUE, FALSE)
  # sum of outliers
  num_outliers <- sum(outliers)

  return(num_outliers)
}

num_outliers <- sapply(df, count_outliers2)
print(num_outliers)
```

```
##              Pregnancies                  Glucose            BloodPressure
##                        4                        0                       14
##            SkinThickness                  Insulin                      BMI
##                       35                       49                        8
## DiabetesPedigreeFunction                      Age                  Outcome
##                       29                        9                        0
```

5. STATISTICAL ANALYSIS

a) Checking Precondition

shapiro-wilk test

```
shapiro.test(df$Glucose)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$Glucose
## W = 0.96986, p-value = 1.733e-11
```

```r
shapiro.test(df$BloodPressure)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$BloodPressure
## W = 0.98782, p-value = 5.255e-06
```

```r
shapiro.test(df$SkinThickness)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$SkinThickness
## W = 0.91367, p-value < 2.2e-16
```

```r
shapiro.test(df$Insulin)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$Insulin
## W = 0.66564, p-value < 2.2e-16
```

```r
shapiro.test(df$BMI)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$BMI
## W = 0.9794, p-value = 6.247e-09
```

```r
shapiro.test(df$DiabetesPedigreeFunction)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$DiabetesPedigreeFunction
## W = 0.83652, p-value < 2.2e-16
```

```r
shapiro.test(df$Outcome)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$Outcome
## W = 0.60251, p-value < 2.2e-16
```

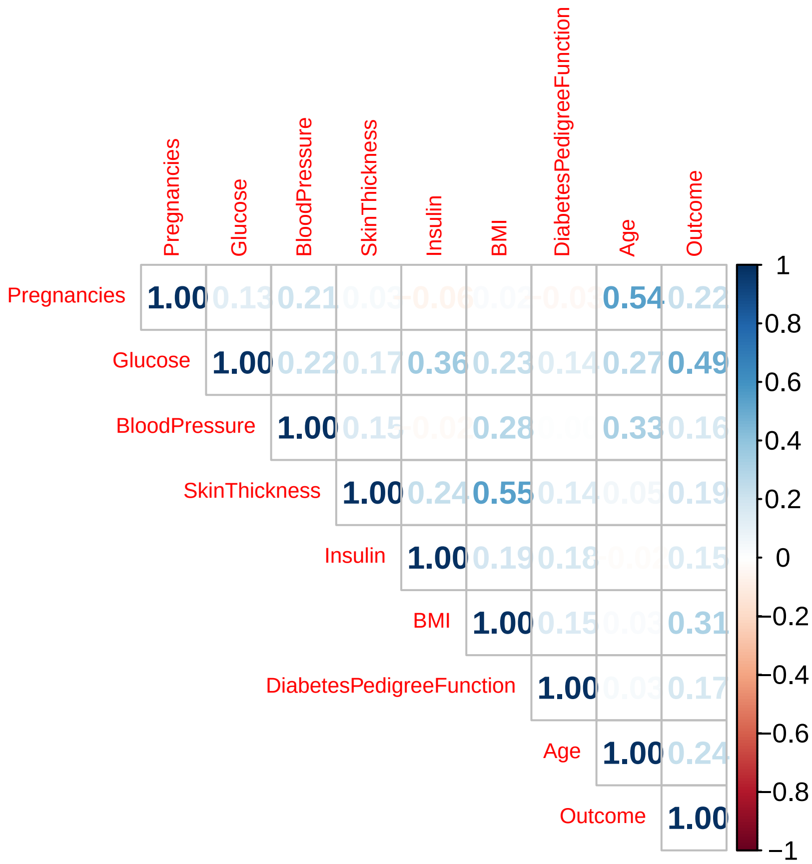
The Shapiro-Wilk tests shows that none of these variables has a normal distribution (mostly because outliers). The low p-values (all < 0.05) indicate strong evidence against the assumption of normality, thus we accept the null hypothesis of non-normality.

Although most of the variables didn't pass the normality test, we are still going to check the relationship between them.

b) Find Correlation

Correlation Plot

```r
c <- cor(df)
corrplot(c, type = "upper", method = "number", tl.cex = 0.7)
```



Here we can see there are variables that have a quite strong relationship:
1. SkinThickness vs Insulin = obesity and the insulin resistant
2. Age vs Pregancies
3. Glucose vs Outcomes = result of OGTT for people with diabetes
4. Glucose vs Insulin
5. BloodPressure vs Age
6. BMI vs Outcome

Pearson's Correlation

```r
correlation1 <- cor.test(df$Glucose, df$Outcome, method = "pearson")
correlation1
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$Glucose and df$Outcome
## t = 15.679, df = 766, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
```

```
##  0.4374171 0.5446602
## sample estimates:
##       cor
## 0.4929084
```

Correlation Coefficient (cor): The test shows a statistically significant positive correlation between Glucose and Outcome, though the strength of the correlation is relatively strong (0.49).

Significance (p-value): The p-value of 2.2e-16 is very small (less than 0.05), suggesting strong evidence against the null hypothesis. Therefore, we reject the null hypothesis that there is no correlation between Glucose and Outcome.

Confidence Interval: We are 95% confident that the true population correlation coefficient falls between 0.4374171 and 0.5446602.

Conclusion:
There is a statistically significant, strong-positive correlation (r = 0.4929084) between Glucose and Outcome. This suggests that as Glucose increases, there tends to be a significant increase in the Outcome. However, the correlation is strong, as the correlation coefficient is relatively close to one. In summary, while there is a statistically significant positive correlation between Glucose and Outcome, and so (r^2 = 0.24295015047576), 24.3% variance of the Outcome can be explained by knowing the Glucose level. So, Glucose alone may not be a strong predictor of Outcome (diabetes). Other factors likely contribute more significantly to determining the outcome.

```r
correlation2 <- cor.test(df$Insulin, df$Outcome, method = "pearson")
correlation2
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$Insulin and df$Outcome
## t = 4.1548, df = 766, p-value = 3.622e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.07853781 0.21692327
## sample estimates:
##       cor
## 0.1484572
```

Correlation Coefficient (cor): The test shows a statistically significant positive correlation between Insulin and Outcome, though the strength of the correlation is relatively weak (0.1484572).

Significance (p-value): The p-value of 3.622e-05 is very small (less than 0.05), suggesting strong evidence against the null hypothesis. Therefore, we reject the null hypothesis that there is no correlation between Insulin and Outcome.

Confidence Interval: We are 95% confident that the true population correlation coefficient falls between 0.07853781 and 0.21692327.

Conclusion:
There is a statistically significant, weak-positive correlation (r = 0.1484572) between Insulin and Outcome. This suggests that as Insulin increases, there tends to be a slight increase in the Outcome. However, the correlation is weak, as the correlation coefficient is relatively close to zero. In summary, while there is a statistically weak positive correlation between Insulin and Outcome, and so (r^2 = 0.02204118108884), only 2.2% variance of the Outcome can be explained by knowing the Insulin level. So, Insulin alone may not be a strong predictor of Outcome (diabetes). Other factors likely contribute more significantly to determining the outcome.

```r
correlation3 <- cor.test(df$BMI, df$Outcome, method = "pearson")
correlation3
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$BMI and df$Outcome
## t = 9.097, df = 766, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2469654 0.3747208
## sample estimates:
##       cor
## 0.3122541
```

Correlation Coefficient (cor): The test shows a statistically significant positive correlation between BMI and Outcome, though the strength of the correlation is relatively weak (0.3122541).

Significance (p-value): The p-value of 2.2e-16 is very small (less than 0.05), suggesting strong evidence against the null hypothesis. Therefore, we reject the null hypothesis that there is no correlation between BMI and Outcome.

Confidence Interval: We are 95% confident that the true population correlation coefficient falls between 0.2469654 and 0.3747208

Conclusion:
There is a statistically significant, weak-positive correlation ($r = 0.3122541$) between BMI and Outcome. This suggests that as BMI increases, there tends to be a slight increase in the Outcome. However, the correlation is weak, as the correlation coefficient is relatively close to zero. In summary, while there is a statistically weak positive correlation between BMI and Outcome, and so ($r\hat{}2 = 0.09757658500881$), only 9.8% variance of the Outcome can be explained by knowing the BMI level. So, BMI alone may not be a strong predictor of Outcome (diabetes). Other factors likely contribute more significantly to determining the outcome.

```r
correlation4 <- cor.test(df$DiabetesPedigreeFunction, df$Outcome, method = "pearson")
correlation4
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$DiabetesPedigreeFunction and df$Outcome
## t = 4.8858, df = 766, p-value = 1.255e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1043836 0.2416168
## sample estimates:
##       cor
## 0.1738441
```

Correlation Coefficient (cor): The test shows a statistically significant positive correlation between DiabetesPedigreeFunction and Outcome, though the strength of the correlation is relatively weak (0.174).

Significance (p-value): The p-value of 1.255e-06 is very small (less than 0.05), suggesting strong evidence against the null hypothesis. Therefore, we reject the null hypothesis that there is no correlation between DiabetesPedigreeFunction and Outcome.

Confidence Interval: This interval indicates that we are 95% confident that the true population correlation coefficient falls between 0.1043836 and 0.2416168.

Conclusion:
There is a statistically significant, weak-positive correlation (r = 0.1738441) between DiabetesPedigreeFunction and Outcome. This suggests that as DiabetesPedigreeFunction increases, there tends to be a slight increase in the Outcome. However, the correlation is not strong, as the correlation coefficient is relatively close to zero. In summary, while there is a statistically weak positive correlation between DiabetesPedigreeFunction and Outcome, and so (r^2 = 0.03023773468), only 3% variance of the Outcome can be explained by knowing the DiabetesPedigreeFunction level. So, DiabetesPedigreeFunction alone may not be a strong predictor of Outcome (diabetes). Other factors likely contribute more significantly to determining the outcome.

```
correlation5 <- cor.test(df$Outcome, df$BloodPressure, method = "pearson")
correlation5
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$Outcome and df$BloodPressure
## t = 4.5721, df = 766, p-value = 5.63e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.0933180 0.2310663
## sample estimates:
##       cor
## 0.1629863
```

Correlation Coefficient (cor): The test shows a statistically significant positive correlation between BloodPressure and Outcome, though the strength of the correlation is relatively weak (0.1629863).

Significance (p-value): The p-value of 1.255e-06 is very small (less than 0.05), suggesting strong evidence against the null hypothesis. Therefore, we reject the null hypothesis that there is no correlation between BloodPressure and Outcome.

Confidence Interval: This interval indicates that we are 95% confident that the true population correlation coefficient falls between 0.0933180 and 0.2310663.

Conclusion:
There is a statistically significant, weak-positive correlation (r = 0.1629863) between BloodPressure and Outcome. This suggests that as BloodPressure increases, there tends to be a slight increase in the Outcome. However, the correlation is not strong, as the correlation coefficient is relatively close to zero. In summary, while there is a statistically weak positive correlation between BloodPressure and Outcome, and so (r^2 = 0.02654817969), only 2.7% variance of the BloodPressure. can be explained by knowing the BloodPressure level. So, Outcome alone may not be a strong predictor of BloodPressure. Other factors likely contribute more significantly to determining the outcome.

c) Check the Regression

```
#      dependent(response) independent
regre <- lm(Outcome ~ DiabetesPedigreeFunction, data =df)
summary(regre)
```

```
##
## Call:
## lm(formula = Outcome ~ DiabetesPedigreeFunction, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8137 -0.3375 -0.2849  0.5963  0.7471
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.23087    0.02953   7.819 1.76e-14 ***
## DiabetesPedigreeFunction  0.25025    0.05122   4.886 1.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.47 on 766 degrees of freedom
## Multiple R-squared:  0.03022,    Adjusted R-squared:  0.02896
## F-statistic: 23.87 on 1 and 766 DF,  p-value: 1.255e-06
```

6. DISCUSSION

From the Pima Indians sample that took OGTT we gained knowledge, that there are two factors that influence diabetes.

OGTT results show the insulin resistance one person has, by looking at Glucose and Insulin levels after 2 hours (most effective). If a person has high glucose results, this indicates insulin failure in breaking down glucose into smaller substances for use by body cells. High insulin results also indicate that the insulin in the body has a delayed response and fails to break down glucose (the pancreas only produces insulin, but the insulin does not work properly).

Firstly, lifestyle took major parts in developing diabetes. People with higher BMI and SkinThickness tends to suffer from diabetes. People who have high body fat will also have high skin fold thickness. Most people that have obesity, tend to eat unhealthy and unbalanced food, consuming too much sugar and processed food. As a result, they are more likely to develop diabetes. Most people underestimate a healthy lifestyle and think that it is not important. Many people from a young age do not pay attention to the physical activity, eventhough the negative impacts can be felt from a young age, it is proven that diabetes begins to develop from the age of 35, and continues to develop in line with increasing age.

Second, heredity factor. People who have a history of diabetes in their family are more likely to develop diabetes. People who have a high DiabetesPedigreeFunction have more chance to grow probability of diabetes even with healthy lifestyle.

Why maintaining diabets is important? Diabetes can affect a person's blood pressure due to insulin resistance. This can cause worse complications such as inflammation and vascular damage.

7. CONCLUSION

From the analysis that we have conduct, We conclude that someone can get diabetes because of their daily lifestyle habits. Data have shown that people that have BMI levels of Overweight and Obesity are more likely to have diabetes. That's why it's really important to maintain a healthy lifestyle every day. However, there's also a chance that we can get diabetes due to family history. Genetic factor can increase the likelihood of developing diabetes, even with a healthy lifestyle. So, it's very important to check if we carry diabetes or not and let's also always maintain your glucose blood in normal level for avoiding any further complication.

For future research, I suggest looking deeper at lifestyle factors and their influence on diabetes, so that in the future there can be an efficient program to prevent diabetes from developing in people who already have hereditary factors.

8. REFFERENCES

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4418458/
https://towardsdatascience.com/pima-indian-diabetes-prediction-7573698bd5fe