PREVALENCE OF DIABETES AMONG PEOPLE LIVING WITH HIV IN PUNE, INDIA

INTRODUCTION

This data analysis report investigates the prevalence of diabetes among individuals living with HIV in Pune, Maharashtra, India. The study aims at understanding the co-occurrence of diabetes mellitus and the Human Immunodeficiency Virus(HIV), and its implications on public health management . Data was collected from medical records of patients in Sassoon Hospital in Pune, which only included the individuals who were diagnosed with HIV. This is a retrospective analysis of patient records that includes demographic, socio – economic and clinical information to determine the prevalence of diabetes within the HIV-positive population.

The report discusses the coexistence of diabetes mellitus and the Human Immunodeficiency Virus(HIV) on the overall health and management of People Living with HIV(PLHIV) in Pune, highlighting the need for integrated healthcare strategies and targeted interventions to address both conditions effectively. Using Logistic Regression and Forestplots, we tracked down major risk factors that heavily contributed towards developing diabetes among PLHIV.

HIV has been a major public health concern for several decades, affecting millions of individuals around the world. Similarly, over 420 million people worldwide are affected by diabetes mellitus, a metabolic disease characterized by high blood-glucose levels. The co-existence of HIV and diabetes presents a critical health challenge in today's world.

There is no cure for HIV infection, but with an effective diagnosis and treatment, significant progress has been made in managing HIV and extending the lives of those affected. An emerging concern is the increased prevalence of diabetes in individuals living with HIV which is between 2% - 14%.

The study population includes 521 adult patients diagnosed with HIV who received ART between 2015 and 2019. A total of 44 (8.4%) patients were found to have diabetes*. Out of 233 males, 27 (11.6%) had diabetes while, out of 283 females, 17 (6%) were diabetic.

Initially, to understand the data, an Exploratory Data Analysis (EDA) was carried out with contingency tables and boxplots. Other statistical tools such as multiple logistic regression, and data visualization tools like forestplots were used to analyse the data. Microsoft Excel and R-Software were extensively used throughout the analysis.

This report mainly consists of three tables showing the prevalence of diabetes in PLHIV through various variables in the whole dataset, in males and in females. The report comprises of all methods used, result, conclusions, and recommendations.

EXPLORATORY DATA ANALYSIS

DATA OVERVIEW :

Characteristic	Median	Inter-Quartile Range
Age	42	36 - 48
Height	158	153 – 166
Weight	53	45 - 61
Mid – Upper Arm Circumference	27	24 - 29
Waist to Hip Ratio	0.89	0.82 – 0.94
Total Cholesterol	165	142 – 192
HDL	43	35 – 53
LDL	95	77-115
Triglycerides	103	73 – 151

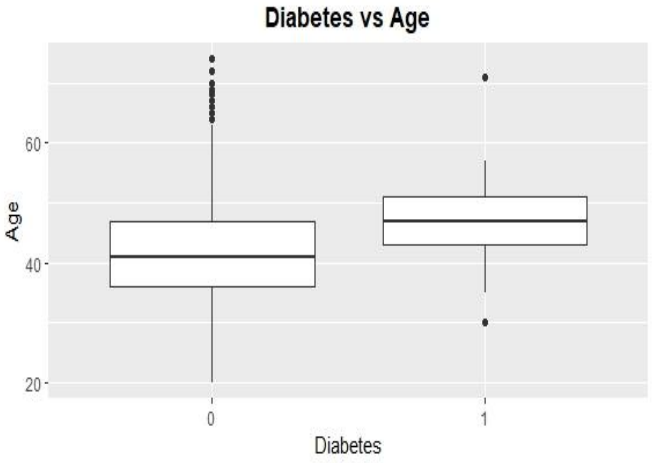
This data was collected at the Sassoon hospital. The dataset dimensions are 521 x 252 where the 252 variables are essentially the questions asked to the HIV patients regarding their demographic, socio-economic and clinical information.

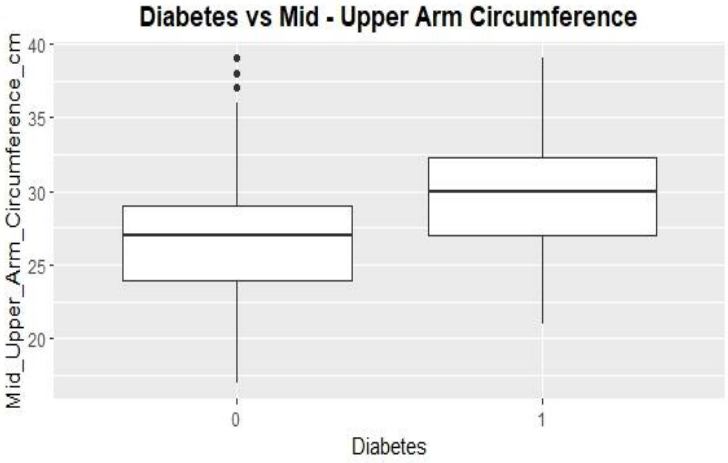
Due to the lacking data, some of the variables in this study were combined and then used. Some variables were converted to a binary scale. These variables and their rules are written below :

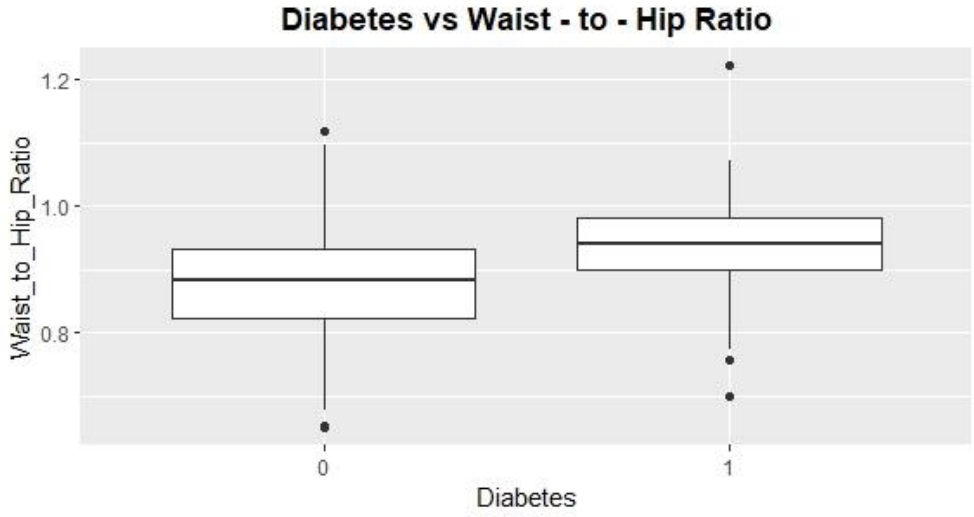
- **Diabetes :**
  - The answer YES to "Have you ever been told by a doctor or other health care worker that you have raised blood sugar or diabetes?" OR
  - The value of HBA1C greater than 6.5% OR
  - The Blood Glucose Level greater than 200 mg/dL
- **Total Cholesterol :**
  - Value greater than 200 is assigned to 1
  - Value lesser than 200 is assigned to 0
- **HDL :**
  - Value lesser than 50 is assigned to 1
  - Value greater than 50 is assigned to 0
- **LDL :**
  - Value greater than 130 is assigned to 1
  - Value lesser than 130 is assigned to 0
- **Triglycerides :**
  - Value greater than 300 is assigned to 1
  - Value lesser than 300 is assigned to 0
- **Obese :**
  - BMI greater than 25 is assigned to 1
  - BMI lesser than 25 is assigned to 0
- **Rigorous Activity :**
  - The answer YES to the question "Does your work involve moderate-intensity activity that causes small increases in breathing or heart rate like (brisk walking, carrying, or lifting light loads) for at least 10 minutes continuously?" is assigned to 1
  - The answer NO is assigned to 0

BOXPLOTS :

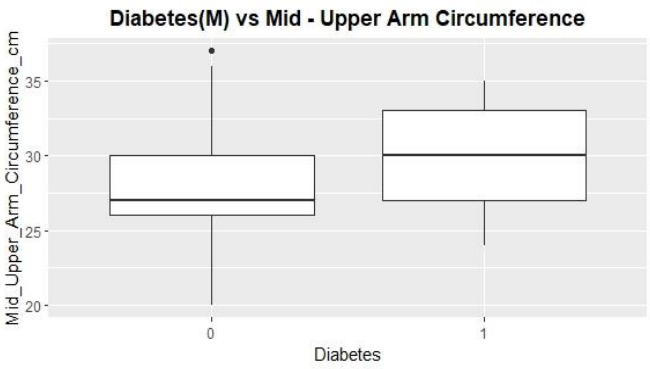
- **Whole Dataset**

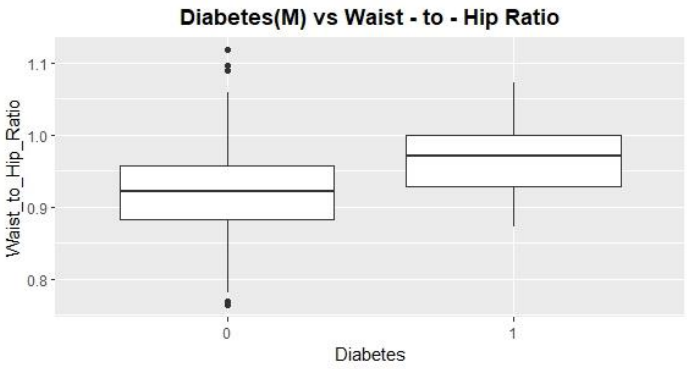




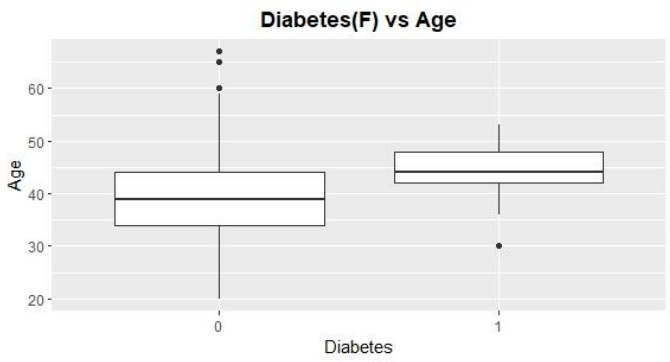


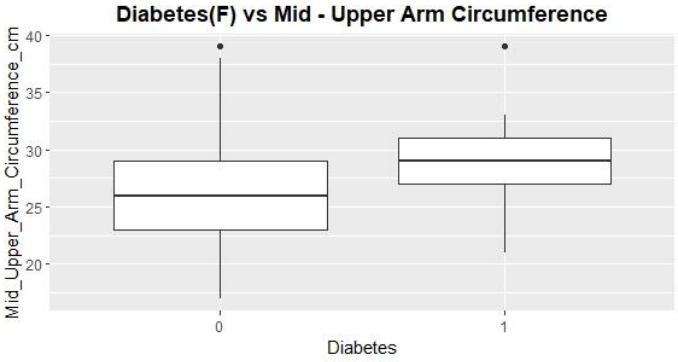
- **Male Subset**

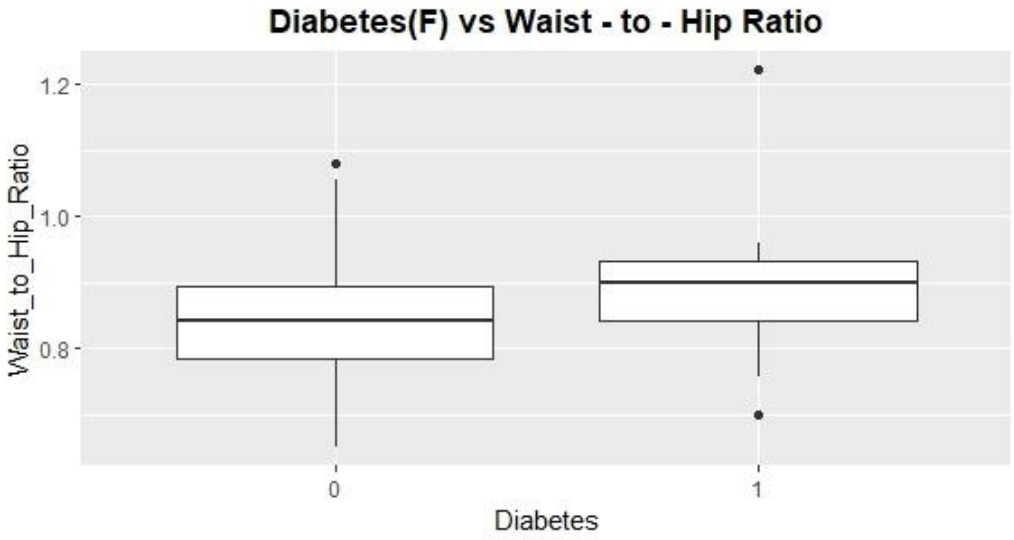




➢ Female Subset







INFERENCES :

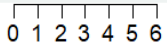
The boxplots above indicate that overall *Age* is a significant characteristic in determining diabetes. But, when we look at the Male and Female subset we find that *Age* is a major factor in Females. The variables *Mid-upper-arm circumference* and *Waist-to-Hip Ratio* are the variables that shows whether or not the individual is obese. These variables are significantly higher in diabetic individuals in both Males and Females which indicates the positive correlation between obesity and diabetes in PLHIV.

These graphs also suggest that, measuring these variables can help identify diabetes among PLHIV at an early stage and an alternative treatment for HIV could be given.

FORESTPLOTS

❖ WHOLE DATASET

Characteristic	Prevalence of Diabetes (N = 44)	OR (95% CI) Univariate	OR (95% CI) Multivariate	Forest Plot	p-value*
Gender					
Male	27 (11%)	2.05 (1.09, 3.86)	1.08 (0.52, 2.24)		0.84
Female	17 (6%)	Ref.	Ref.		
Age (years)					
>40	39 (12%)	5.38 (2.08, 13.89)	3.14 (1.34, 8.13)		0.01
<40	5 (2%)	Ref.	Ref.		
Rigorous Activity					
No	20 (13%)	2.15 (1.15, 4.02)	1.91 (0.92, 3.95)		0.08
Yes	24 (7%)	Ref.	Ref.		
Hypertension					
Yes	34 (13%)	3.5 (1.69, 7.25)	2.07 (1.05, 4.15)		0.04
No	10 (4%)	Ref.	Ref.		
Raised BP					
Yes	27 (13%)	2.51 (1.33, 4.73)	0.74 (0.32, 1.8)		0.5
No	17 (5%)	Ref.	Ref.		
Obese					
Yes	23 (18%)	4.08(2.17, 7.66)	2.63 (1.27, 5.48)		0.01
No	21 (5%)	Ref.	Ref.		
Total Cholesterol					
<200 (desired)	30 (7%)	2.22 (1.13, 4.36)	0.44 (0.13, 1.35)		0.16
>200	14 (14%)	Ref.	Ref.		
LDL					
>130	10 (15%)	2.17 (1.02, 4.62)	2.51 (0.75, 8.06)		0.13
<130 (desired)	34 (7%)	Ref.	Ref.		
Triglycerides					
>300	13 (43%)	11.35 (5.06, 25.47)	11.52 (4.2, 32.28)		<0.01
<300 (desired)	31 (6%)	Ref.	Ref.		

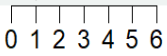


➢ Inferences :

- Age : We observe that PLHIV who are aged 40 years or older are ~3 times more likely to have diabetes than the people who are less than 40 years old.
- Hypertension : Out of all, the PLHIV who have hypertension are ~2 times more likely to have diabetes as compared to the PLHIV who don't have hypertension.
- Obese : The PLHIV who are obese (BMI > 25) are ~2.5 times more likely to have diabetes with respect to the people who are not obese.
- Triglycerides : The PLHIV who have triglycerides greater than 300mg/dL are ~11.5 times more likely to have diabetes than the people who have triglycerides less than 300 mg/dL.

❖ MALES

Characteristic	Prevalence of Diabetes (N = 27)	OR (95% CI) Univariate	OR (95% CI) Multivariate	Forest Plot	p-value*
Age					
>40	24 (14%)	2.23 (0.81, 6.16)	2.19 (0.73, 7.7)		0.19
<40	3 (5%)	Ref.	Ref.		
Obese					
Yes	17 (21%)	3.86 (1.67, 8.9)	3.6 (1.52, 8.87)		<0.01
No	10 (6%)	Ref.	Ref.		
Total Cholesterol					
<200 (desired)	16 (9%)	2.61 (1.13, 6.02)	1.13 (0.37, 3.15)		0.81
>200	11 (20%)	Ref.	Ref.		
Triglycerides					
>300	8 (40%)	6.81 (2.48, 18.71)	5.42 (1.59, 18.69)		<0.01
<300 (desired)	19 (9%)	Ref.	Ref.		

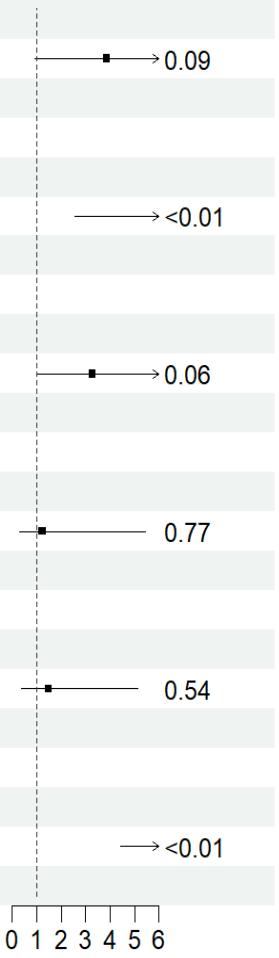


➢ Inferences :

- Age : PLHIV males who are aged above 40 are ~2 times more likely to have diabetes than the males who are less than 40 years old.
- Obese : The PLHIV males who are obese (BMI > 25) are ~3.5 times more likely to have diabetes with respect to the people who are not obese.
- Triglycerides : The PLHIV males who have triglycerides greater than 300mg/dL are ~5.5 times more likely to have diabetes than the people who have triglycerides less than 300 mg/dL.

❖ FEMALES

Characteristic	Prevalence of Diabetes (N = 17)	OR (95% CI) Univariate	OR (95% CI) Multivariate	Forest Plot	p-value*
Age (years)					
>40	15 (10%)	6.09 (1.68, 22.16)	3.86 (0.93, 24.43)		0.09
<40	2 (1%)	Ref.	Ref.		
Rigorous Activity					
No	10 (14%)	4.9 (1.79, 13.44)	8.84 (2.56, 36.36)		<0.01
Yes	7 (3%)	Ref.	Ref.		
Hypertension					
Yes	14 (10%)	3.64 (1.48, 8.94)	3.28 (0.99, 12.18)		0.06
No	3 (2%)	Ref.	Ref.		
Raised BP					
Yes	10 (11%)	3.44 (1.27, 9.37)	1.22 (0.31, 5.44)		0.77
No	7 (4%)	Ref.	Ref.		
Obese					
Yes	6 (14%)	3.27 (1.14, 9.37)	1.49 (0.39, 5.11)		0.54
No	11 (5%)	Ref.	Ref.		
Triglycerides					
>300	5 (50%)	21.75 (5.54, 85.42)	23.35 (4.42, 144)		<0.01
<300 (desired)	12 (4%)	Ref.	Ref.		



➢ Inferences :

- Rigorous Activity : We observe that PLHIV females who are not engaged in a rigorous physical activity are ~9 times more likely to have diabetes than the people who are engaged in such activities.
- Triglycerides : The PLHIV females who have triglycerides greater than 300mg/dL are ~23 times more likely to have diabetes than the people who have triglycerides less than 300 mg/dL.

NOTE : The * indicated in these 3 tables above signifies that the p-values obtained are through the multivariate logistic model.

APPENDIX (R-CODE)

```r
#Importing Libraries
library(readxl)
library(epiDisplay)
library(lmtest)
library(MASS)
library(car)
library(sigmoid)
library(multcomp)
library(ResourceSelection)
library(boot)
library(ggplot2)
library(vcd)
library(caTools)
library(caret)
library(tidyverse)
library(forestploter)
library(grid)
library(epitools)


#Importing Dataset
ncd <- read_excel("C:/Users/dhruv/OneDrive/Desktop/Project Nikhil Sir/Prevalenc
e of Diabetes in PLHIV/Data/NCD_Database_27_02_19.xlsx")

View(ncd)
summary(ncd)


#For splitting data according to gender
ncdm = ncd %>% filter(Sex == "M")
ncdf = ncd %>% filter(Sex == "F")


#Creating variable : doc
doc = c()
for (i in 1:length(ncd$`Have you ever been told by a doctor or other health car
e worker that you have raised blood sugar or diabetes`)) {
  if (is.na(ncd$`Have you ever been told by a doctor or other health care worke
r that you have raised blood sugar or diabetes`[i])) {
    doc = append(doc, "Missing value")
  } else if (ncd$`Have you ever been told by a doctor or other health care work
er that you have raised blood sugar or diabetes`[i] == "Yes") {
    doc = append(doc, "Yes")
  } else {
    doc = append(doc, "No")
  }
}

ncd$doc = doc
```

```r
#Creating variable : hba1c
hba1c = c()
for (i in 1:length(ncd$`Please enter HbA1c value in %`)) {
  if (is.na(ncd$`Please enter HbA1c value in %`[i])) {
    hba1c = append(hba1c, 0)
  }
  else {
    hba1c = append(hba1c, ncd$`Please enter HbA1c value in %`[i])
  }
}

ncd$hba1c = hba1c


#Creating variable : glucose
glucose = c()
for (i in 1:length(ncd$Glucose)) {
  if (is.na(ncd$Glucose[i])) {
    glucose = append(glucose, 0)
  }
  else {
    glucose = append(glucose, ncd$Glucose[i])
  }
}

ncd$glucose = glucose


#Defining the response variable : Diabetes (db)
db = c()
for (i in 1:521) {
  if (is.na(ncd$hba1c[i])){
    na.rm = TRUE
  } else if (ncd$hba1c[i] >= 6.5) {
    db = append(db, 1)
  } else if (ncd$glucose[i] >= 200) {
    db = append(db, 1)
  } else if (ncd$doc[i] == "Yes") {
    db = append(db, 1)
  } else {
    db = append(db, 0)
  }
}

ncd$diabetic = db;ncd$diabetic
```

```r
#Creating variable : Tuberculosis
tb_d = c()
for (i in 1:521){
  if ("Yes" %in% ncd$`Have you ever been told that you have had TB?`[i]){
    tb_d = append(tb_d, 1)
  }
  else if ("Yes" %in% ncd$`Were you put on TB therapy?`[i]){
    tb_d = append(tb_d, 1)
  }
  else{
    tb_d = append(tb_d, 0)
  }
}
ncd$tb_d = tb_d


#Creating variable : raised BP
r_bp = c()
for (i in 1:521){
  if (ncd$avg_sys1[i] > 120){
    r_bp = append(r_bp, 1)
  }
  else if ("Yes" %in% ncd$`Have you ever been told by a doctor or other health
care worker that you have raised blood pressure or hypertension?`[i]){
    r_bp = append(r_bp, 1)
  }
  else{
    r_bp = append(r_bp, 0)
  }
}
r_bp


#Creating variable : cooking fuel
cook_fuel = c()
for (i in 1:521){
  if ("Yes" %in% ncd$`Gas Stove`[i]){
    cook_fuel = append(cook_fuel, 1)
  }
  else if ("Yes" %in% ncd$`Electric Stove`[i]){
    cook_fuel = append(cook_fuel, 1)
  }
  else{
    cook_fuel = append(cook_fuel, 0)
  }
}

ncd$cook_fuel = cook_fuel
```

```r
#Creating variable : Cholesterol (cho)
cho = c()
for (i in 1:521){
  if (!is.na(ncd$`Total cholesterol`[i]) && ncd$`Total cholesterol`[i] > 200){
    cho = append(cho, 1)
  }
  else if (is.na(ncd$`Total cholesterol`[i])){
    cho = append(cho, 0)
  }
  else{
    cho = append(cho, 0)
  }
}
cho


#Creating variable : HDL (hdl)
hdl = c()
for (i in 1:521){
  if (!is.na(ncd$HDL[i]) && ncd$HDL[i] < 50){
    hdl = append(hdl, 1)
  }
  else if (is.na(ncd$HDL[i])){
    hdl = append(hdl, 0)
  }
  else{
    hdl = append(hdl, 0)
  }
}
hdl


#Creating variable : LDL (ldl)
ldl = c()
for (i in 1:521){
  if (!is.na(ncd$`LDL (calculated)`[i]) && ncd$`LDL (calculated)`[i] > 130){
    ldl = append(ldl, 1)
  }
  else if (is.na(ncd$`LDL (calculated)`[i])){
    ldl = append(ldl, 0)
  }
  else{
    ldl = append(ldl, 0)
  }
}
ldl
```

```r
#Creating variable : Triglycerides (trigl)
trigl = c()
for (i in 1:521){
  if (!is.na(ncd$Triglycerides[i]) && ncd$Triglycerides[i] > 300){
    trigl = append(trigl, 1)
  }
  else if (is.na(ncd$Triglycerides[i])){
    trigl = append(trigl, 0)
  }
  else{
    trigl = append(trigl, 0)
  }
}
ncd$trigl = trigl


#Creating variable : Age above 40 years (age40)
age40 = c()
for (i in 1:521){
  if (!is.na(ncd$`Age(Years)`[i]) && ncd$`Age(Years)`[i] >= 40) {
    age40 = append(age40, 1)
  }
  else if (is.na(ncd$`Age(Years)`[i])){
    age40 = append(age40, 2)
  }
  else{
    age40 = append(age40, 0)
  }
}
ncd$age40 = age40


#Creating variable : Income (inc)
inc = c()
for (i in 1:521){
  if (is.na(ncd$`Monthly household income (in INR)`[i])){
    inc = append(inc, 0)
  }
  else if (ncd$`Monthly household income (in INR)`[i] == "1 000 - 4 999" || ncd
$`Monthly household income (in INR)`[i] == "5 000 - 9 999") {
    inc = append(inc, 1)
  }
  else{
    inc = append(inc, 0)
  }
}
ncd$inc = inc
```

```r
#Creating variable : Obese (obes)
obes = c()
for (i in 1:521){
  if (is.na(ncd$BMI[i])){
    obes = append(obes, 0)
  }
  else if (ncd$BMI[i] > 25) {
    obes = append(obes, 1)
  }
  else{
    obes = append(obes, 0)
  }
}
ncd$obes = obes



#For obtaining the p-value, Odds Ratio, 95%CI of Odds Ratio

#model = glm(diabetic ~ <add variable name>, data = ncd, family = "binomial")
#summary(model)
#logistic.display(model)


#2x2 contingency tables
ncd$diabetic = ordered(ncd$diabetic,levels = c(0,1),labels = c("No", "Yes"))
smoked = ncd$`Have you ever smoked?`
hyp = ncd$`hyper 1`
ncd$hyp = ordered(hyp, levels = c(0,1), labels = c("No", "Yes"))

#Demographic Variables
gender = table(db, ncd$Sex);gender
age_40 = table(db, age40);age_40
edu = table(db, ncd$education);edu #or
inco = table(db, inc);inco
loc = table(db, ncd$`Living location`);loc
empl = table(db, ncd$`Are you employed?`);empl
rig_act = table(db, ncd$`Does your work involve moderate -intensity activity th
at causes small increases in breathing or heart rate like (brisk walking    car
rying or lifting light loads) for at least 10 minutes continuously ?`);rig_act
cookfuel = table(db, cook_fuel);cookfuel
hfias = table(db, ncd$`HFIAS Score`);hfias_score = ncd$`HFIAS Score`;hfias


#Social(Lifestyle) Variables
alc = table(db, ncd$`Do you drink alcohol?`);alco = ncd$`Do you drink alcohol?`
;alc
smo = table(db, smoked);smo
hyper = table(db, hyp);hyper
bp = table(db, r_bp);bp
obs = table(db, obes);obs
chol = table(db, cho);ncd$cho = cho;chol
hd_l = table(db, hdl);hd_l
ld_l = table(db, ldl);ld_l
trig = table(db, trigl);ncd$trigl = trigl;trig
```

```r
#Clinical Variables
tuber = table(db, tb_d);tuber
renal = table(db, ncd$`Have you ever been told by a doctor or other health care worker that you have chronic renal disease?`);ncd$renal_d = ncd$`Have you ever been told by a doctor or other health care worker that you have chronic renal disease?`;renal
liver = table(db, ncd$`Have you ever been told by a doctor or other health care worker that you have chronic liver disease?`);ncd$liver_d = ncd$`Have you ever been told by a doctor or other health care worker that you have chronic liver disease?`;liver
tb_inf = table(db, ncd$`QGIT/IGRA Result`);ncd$tb_in = ncd$`QGIT/IGRA Result`;
tb_inf


#BOXPLOTS

#Diabetic vs Height
Diabetes = factor(db)
Height_cm = ncd$`Height (cm)`

df1 <- data.frame(db, ncd$`Height (cm)`)
ggplot(df1, aes(x = Diabetes, y = Height_cm)) +
  geom_boxplot() +
  ggtitle("Diabetes vs Height") +
  theme(plot.title = element_text(face = "bold",hjust = 0.5))

#Diabetic vs Weight
Diabetes = factor(db)
Weight_kg = ncd$`Weight (kg)`

df2 <- data.frame(db, ncd$`Weight (kg)`)
ggplot(df2, aes(x = Diabetes, y = Weight_kg)) +
  geom_boxplot() +
  ggtitle("Diabetes(F) vs Weight") +
  theme(plot.title = element_text(face = "bold",hjust = 0.5))

#Diabetic vs Cholesterol
Diabetes = factor(db)
Total_Cholesterol = ncd$`Total cholesterol`

df3 <- data.frame(db, ncd$`Total cholesterol`)
ggplot(df3, aes(x = Diabetes, y = Total_Cholesterol)) +
  geom_boxplot() +
  ggtitle("Diabetes vs Cholesterol") +
  theme(plot.title = element_text(face = "bold",hjust = 0.5))

#Diabetic vs Age
Diabetes = factor(db)
Age = ncd$`Age(Years)`

df4 <- data.frame(db, ncd$`Age(Years)`)
ggplot(df4, aes(x = Diabetes, y = Age)) +
  geom_boxplot() +
  ggtitle("Diabetes(F) vs Age") +
  theme(plot.title = element_text(face = "bold",hjust = 0.5))
```

```r
#Diabetic vs Mid Upper Arm Circumference
Diabetes = factor(db)
Mid_Upper_Arm_Circumference_cm = ncd$`Mid-upper arm circumference (cm)`

df5 <- data.frame(db, ncd$`Mid-upper arm circumference (cm)`)
ggplot(df5, aes(x = Diabetes, y = Mid_Upper_Arm_Circumference_cm)) +
  geom_boxplot() +
  ggtitle("Diabetes(F) vs Mid - Upper Arm Circumference") +
  theme(plot.title = element_text(face = "bold",hjust = 0.5))

#Diabetic vs Waist to Hip Ratio
Diabetes = factor(db)
Waist_to_Hip_Ratio = ncd$`Waist to hip ratio`

df6 <- data.frame(db, ncd$`Waist to hip ratio`)
ggplot(df6, aes(x = Diabetes, y = Waist_to_Hip_Ratio)) +
  geom_boxplot() +
  ggtitle("Diabetes(F) vs Waist - to - Hip Ratio") +
  theme(plot.title = element_text(face = "bold",hjust = 0.5))



#Model with all variables for the whole dataset

rig = ncd$`Does your work involve moderate -intensity activity that causes smal
l increases in breathing or heart rate like (brisk walking    carrying or lifti
ng light loads) for at least 10 minutes continuously ?`
rig1 = relevel(as.factor(rig), ref = "Yes")
ncd$r_bp = r_bp
ncd$ldl = ldl


model = glm(diabetic ~ Sex + age40 + rig1 + hyp + r_bp + obes + cho + ldl + tri
gl, data = ncd, family = "binomial")
summary(model)

OR = exp(coef(model))
CI_95 = exp(confint(model))
table = cbind(OR, CI_95);table



#Model with all variables for males

modelm = glm(db ~ age40 + obes + cho + trigl, data = ncdm, family = "binomial")
summary(modelm)

OR = exp(coef(modelm))
CI_95 = exp(confint(modelm))
table = cbind(OR, CI_95);table
```

```r
#Model with all variables for females

modelf = glm(db ~ age40 + rig1 + hyp + r_bp + obes + trigl, data = ncdf, family
= "binomial")
summary(modelf)

OR = exp(coef(modelf))
CI_95 = exp(confint(modelf))
table = cbind(OR, CI_95);table



#Forest Plots : Whole Dataset
dt = read_excel("C:\\Users\\dhruv\\OneDrive\\Desktop\\Project Nikhil Sir\\Preva
lence of Diabetes in PLHIV\\Data\\Multivariate.xlsx")
dt$est = ifelse(is.na(dt$est), "",dt$est)
dt$lower = ifelse(is.na(dt$lower), "",dt$lower)
dt$upper = ifelse(is.na(dt$upper), "",dt$upper)
dt$`Prevalence of Diabetes (N = 44)` = ifelse(is.na(dt$`Prevalence of Diabetes
(N = 44)`), "",dt$`Prevalence of Diabetes (N = 44)`)
dt$`OR (95% CI) Univariate` = ifelse(is.na(dt$`OR (95% CI) Univariate`), "",dt$
`OR (95% CI) Univariate`)
dt$`OR (95% CI) Multivariate` = ifelse(is.na(dt$`OR (95% CI) Multivariate`), ""
,dt$`OR (95% CI) Multivariate`)
dt$`p-value*` = ifelse(is.na(dt$`p-value*`), "",dt$`p-value*`)
dt$`Forest Plot` = ifelse(is.na(dt$`Forest Plot`), "",dt$`Forest Plot`)
dt$Characteristic = ifelse(dt$`Prevalence of Diabetes (N = 44)` == "", dt$Chara
cteristic, paste0("   ", dt$Characteristic))


est = as.numeric(dt$est)
lower = as.numeric(dt$lower)
upper = as.numeric(dt$upper)

tm = forest_theme(base_size = 9)


p = forest(dt[,c(1:5,9)], est = est, lower = lower,
           upper = upper,
           sizes = 0.4,
           ci_column = 5,
           ref_line = 1,
           xlim = c(0, 6),
           theme = tm)

p = edit_plot(p, row = c(1, 4, 7, 10, 13, 16, 19, 22, 25),
              gp = gpar(fontface = "bold"))
plot(p)
```

```r
#Forest Plots : Male Dataset
dtm = read_excel("C:\\Users\\dhruv\\OneDrive\\Desktop\\Project Nikhil Sir\\Prev
alence of Diabetes in PLHIV\\Data\\Multivariate_Male.xlsx")
dtm$est = ifelse(is.na(dtm$est), "",dtm$est)
dtm$lower = ifelse(is.na(dtm$lower), "",dtm$lower)
dtm$upper = ifelse(is.na(dtm$upper), "",dtm$upper)
dtm$`Prevalence of Diabetes (N = 27)` = ifelse(is.na(dtm$`Prevalence of Diabete
s (N = 27)`), "",dtm$`Prevalence of Diabetes (N = 27)`)
dtm$`OR (95% CI) Univariate` = ifelse(is.na(dtm$`OR (95% CI) Univariate`), "",d
tm$`OR (95% CI) Univariate`)
dtm$`OR (95% CI) Multivariate` = ifelse(is.na(dtm$`OR (95% CI) Multivariate`),
"",dtm$`OR (95% CI) Multivariate`)
dtm$`p-value*` = ifelse(is.na(dtm$`p-value*`), "",dtm$`p-value*`)
dtm$`Forest Plot` = ifelse(is.na(dtm$`Forest Plot`), "",dtm$`Forest Plot`)
dtm$Characteristic = ifelse(is.na(dtm$Characteristic),"",dtm$Characteristic)
dtm$Characteristic = ifelse(dtm$`Prevalence of Diabetes (N = 27)` == "", dtm$Ch
aracteristic, paste0("   ", dtm$Characteristic))


estm = as.numeric(dtm$est)
lowerm = as.numeric(dtm$lower)
upperm = as.numeric(dtm$upper)


m = forest(dtm[,c(1:5,9)], est = estm, lower = lowerm,
           upper = upperm,
           sizes = 0.4,
           ci_column = 5,
           ref_line = 1,
           xlim = c(0, 6))

m = edit_plot(m, row = c(1, 5, 9, 13),
              gp = gpar(fontface = "bold"))
plot(m)
```

```r
#Forest Plots : Female Dataset
dtf = read_excel("C:\\Users\\dhruv\\OneDrive\\Desktop\\Project Nikhil Sir\\Prev
alence of Diabetes in PLHIV\\Data\\Multivariate_Female.xlsx")
dtf$est = ifelse(is.na(dtf$est), "",dtf$est)
dtf$lower = ifelse(is.na(dtf$lower), "",dtf$lower)
dtf$upper = ifelse(is.na(dtf$upper), "",dtf$upper)
dtf$`Prevalence of Diabetes (N = 17)` = ifelse(is.na(dtf$`Prevalence of Diabete
s (N = 17)`), "",dtf$`Prevalence of Diabetes (N = 17)`)
dtf$`OR (95% CI) Univariate` = ifelse(is.na(dtf$`OR (95% CI) Univariate`), "",d
tf$`OR (95% CI) Univariate`)
dtf$`OR (95% CI) Multivariate` = ifelse(is.na(dtf$`OR (95% CI) Multivariate`),
"",dtf$`OR (95% CI) Multivariate`)
dtf$`p-value*` = ifelse(is.na(dtf$`p-value*`), "",dtf$`p-value*`)
dtf$`Forest Plot` = ifelse(is.na(dtf$`Forest Plot`), "",dtf$`Forest Plot`)
dtf$Characteristic = ifelse(is.na(dtf$Characteristic),"",dtf$Characteristic)
dtf$Characteristic = ifelse(dtf$`Prevalence of Diabetes (N = 17)` == "", dtf$Ch
aracteristic, paste0("   ", dtf$Characteristic))


estf = as.numeric(dtf$est)
lowerf = as.numeric(dtf$lower)
upperf = as.numeric(dtf$upper)


f = forest(dtf[,c(1:5,9)], est = estf, lower = lowerf,
           upper = upperf,
           sizes = 0.4,
           ci_column = 5,
           ref_line = 1,
           xlim = c(0, 6))

f = edit_plot(f, row = c(1, 5, 9, 13, 17, 21),
              gp = gpar(fontface = "bold"))
plot(f)
```