BMI and Hypertension prevalence

**Group Assignment**

## 

|  |  |
| --- | --- |
| NAME | REGISTRATION NUMBER |
| TUKIKO ANYONA | HEM/B/01-04959/2023 |
| NOBERT WAKASALA | HEM/B/01-04973/2023 |
| TIMOTHY ACHALA | HEM/B/01-04976/2023 |
| MARTHA MBITHE | HEM/B/01-04049/2022 |
| SHELLTON CHEPKOECH | HEM/B/01-04961/2023 |
| VALLERY ORWA | HEM/B/01-07752/2023 |

## Study objective

The main objective of the study is to analyze the correlation between BMI and the prevalence of hypertension among adults in different age groups.

## 1. Loading the required libraries

library(tidyverse)  
library(e1071)  
library(psych)  
library(mice)  
library(VIM)  
library(metan)  
library(DescTools)  
library(gtExtras)  
library(gt)  
library(ggcorrplot)

## 2. Importing the data set

Hypertension <- read.csv("E:/data sets/Hypertension.csv")

# Viewing the first 5 variables  
Hypertension %>%   
 head(5) %>%   
 gt() %>%   
 gt\_theme\_guardian() %>%   
 tab\_header(title = "Hypertension first characters")

Table : Hypertension first characters

| gender | age | sysBP | diaBP | BMI | heartRate | Hypertensionrisk |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 39 | 106.0 | 70 | 26.97 | 80 | 0 |
| 0 | 46 | 121.0 | 81 | 28.73 | 95 | 0 |
| 1 | 48 | 127.5 | 80 | 25.34 | 75 | 0 |
| 0 | 61 | 150.0 | 95 | 28.58 | 65 | 1 |
| 0 | 46 | 130.0 | 84 | 23.10 | 85 | 0 |

# Viewing the last 5 variables  
Hypertension %>%   
 tail(5) %>%   
 gt() %>%   
 gt\_theme\_guardian() %>%   
 tab\_header(title = "Hypertension last characters")

Table : Hypertension last characters

| gender | age | sysBP | diaBP | BMI | heartRate | Hypertensionrisk |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 48 | 131.0 | 72 | 22.00 | 84 | 0 |
| 0 | 44 | 126.5 | 87 | 19.16 | 86 | 0 |
| 0 | 52 | 133.5 | 83 | 21.47 | 80 | 0 |
| 1 | 40 | 141.0 | 98 | 25.60 | 67 | 1 |
| 0 | 39 | 133.0 | 86 | 20.91 | 85 | 0 |

# data types  
str(Hypertension)

'data.frame': 4240 obs. of 7 variables:  
 $ gender : int 1 0 1 0 0 0 0 0 1 1 ...  
 $ age : int 39 46 48 61 46 43 63 45 52 43 ...  
 $ sysBP : num 106 121 128 150 130 ...  
 $ diaBP : num 70 81 80 95 84 110 71 71 89 107 ...  
 $ BMI : num 27 28.7 25.3 28.6 23.1 ...  
 $ heartRate : int 80 95 75 65 85 77 60 79 76 93 ...  
 $ Hypertensionrisk: int 0 0 0 1 0 1 0 0 1 1 ...

### Adding age group variable to the entire data set

Hypertension <- Hypertension %>%   
 mutate(age\_group = case\_when(  
 age >= 0 & age < 17.9 ~ "Child",  
 age >= 18 & age < 45.9 ~ "Adult",  
 age >= 46 & age < 59.9 ~ "Middle aged",  
 TRUE ~ "Senior"  
 ))

### Adding BMI category to the entire data set

Hypertension <- Hypertension %>%   
 mutate(bmi\_group = case\_when(  
 BMI < 18.5 ~ "Underweight",  
 BMI >= 18.5 & BMI <= 24.9 ~ "Normalweight",  
 BMI >= 25 & BMI <= 29.9 ~ "Overweight",  
 TRUE ~ "Obese"  
 ))

### Adding other columns (agegroup and bmigroup)

Hypertension <-Hypertension %>%   
 mutate(agegroup= case\_when(age\_group == "Child" ~ "0",  
 age\_group == "Adult" ~ "1",  
 age\_group == "Middle aged" ~ "2",  
 TRUE ~ "3")) %>%   
 mutate(bmigroup = case\_when(bmi\_group == "Underweight" ~ "0",  
 bmi\_group == "Normalweight" ~ "1",  
 bmi\_group == "Overweight" ~ "2",  
 TRUE ~ "3"))

Hypertension<-Hypertension %>%   
mutate(Hypertension\_risk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes")))

### Preparing for plots

theme\_set(theme\_bw()+  
 theme(title = element\_text(color = "#4CBB17",  
 size = 18,  
 face = "bold"),  
 axis.text =   
 element\_text(size = 10,  
 color = "#6495ED",  
 face = "bold"),  
 axis.title = element\_text(size = 12,  
 face = "bold",  
 colour = "#FF5733")))

## 3. Data cleaning

### Calculating the percentage of missing values

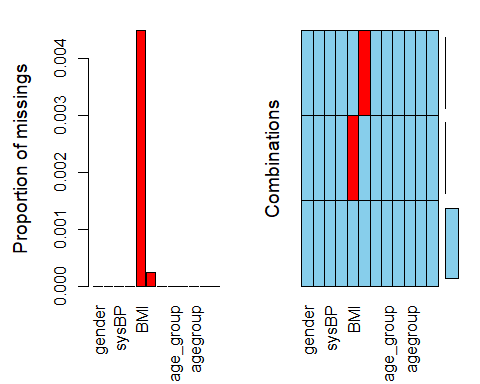
p <- function(Hypertension) {sum(is.na(Hypertension))/length(Hypertension)\* 100}  
apply(Hypertension,2,p)

gender age sysBP diaBP   
 0.00000000 0.00000000 0.00000000 0.00000000   
 BMI heartRate Hypertensionrisk age\_group   
 0.44811321 0.02358491 0.00000000 0.00000000   
 bmi\_group agegroup bmigroup Hypertension\_risk   
 0.00000000 0.00000000 0.00000000 0.00000000

It can be evidently seen that 0.448% of BMI has missing values

### Dealing with the missing data

# visualizing the proportion of missing values  
aggr(Hypertension)



# applied predictive mean matching (pmm) for BMI   
new\_Hypertension <-mice(Hypertension, seed = 300)

iter imp variable  
 1 1 BMI heartRate  
 1 2 BMI heartRate  
 1 3 BMI heartRate  
 1 4 BMI heartRate  
 1 5 BMI heartRate  
 2 1 BMI heartRate  
 2 2 BMI heartRate  
 2 3 BMI heartRate  
 2 4 BMI heartRate  
 2 5 BMI heartRate  
 3 1 BMI heartRate  
 3 2 BMI heartRate  
 3 3 BMI heartRate  
 3 4 BMI heartRate  
 3 5 BMI heartRate  
 4 1 BMI heartRate  
 4 2 BMI heartRate  
 4 3 BMI heartRate  
 4 4 BMI heartRate  
 4 5 BMI heartRate  
 5 1 BMI heartRate  
 5 2 BMI heartRate  
 5 3 BMI heartRate  
 5 4 BMI heartRate  
 5 5 BMI heartRate

# Getting the structure of the imputed data set  
attributes(new\_Hypertension)

$names  
 [1] "data" "imp" "m" "where"   
 [5] "blocks" "call" "nmis" "method"   
 [9] "predictorMatrix" "visitSequence" "formulas" "post"   
[13] "blots" "ignore" "seed" "iteration"   
[17] "lastSeedValue" "chainMean" "chainVar" "loggedEvents"   
[21] "version" "date"   
  
$class  
[1] "mids"

### Getting the new complete data set

# getting the new complete dataset  
Hypertensionn <- complete(new\_Hypertension)

### Changing the bmi\_group and age\_group to factor

Hypertensionn$bmi\_group <- as.factor(Hypertensionn$bmi\_group)

Hypertensionn$age\_group <- as.factor(Hypertensionn$age\_group)

### Creating levels in the BMI variable

Hypertensionn$bmi\_group<- factor((Hypertensionn$bmi\_group), levels = c("Obese",  
 "Overweight",  
 "Normalweight",  
 "Underweight"))

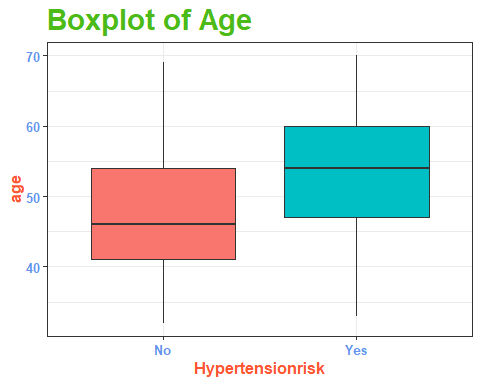
### Creating levels in the age group variable

Hypertensionn$age\_group<- factor((Hypertensionn$age\_group), levels = c("Child",  
 "Adult",  
 "MiddleAged",  
 "Senior",  
 "Unknown"))

### Dealing with outliers

#### We will check for outliers in the age variable

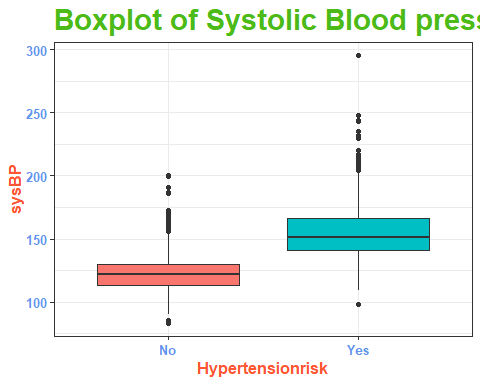
Hypertensionn %>%  
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%   
 ggplot(aes(Hypertensionrisk,age))+  
 geom\_boxplot(aes(fill = Hypertensionrisk ),show.legend = F)+  
 labs(title = "Boxplot of Age")



It can be evidently seen as per the plot that age has NO outliers

#### Checking for outliers in the systolic blood pressure variable

Hypertensionn %>%  
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%   
 ggplot(aes(Hypertensionrisk,sysBP))+  
 geom\_boxplot(aes(fill = Hypertensionrisk ),show.legend = F)+  
 labs(title = "Boxplot of Systolic Blood pressure")

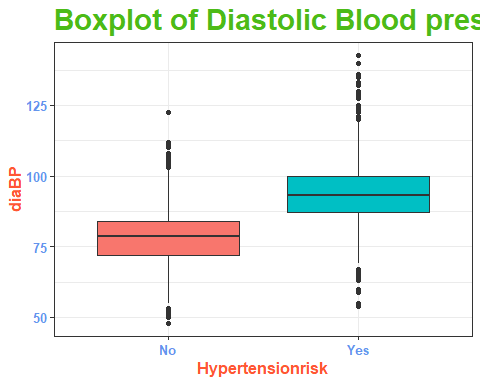


It can be seen that systolic blood pressure has outliers, we will winsorize the outliers

# winsorizing the systolic blood pressure in order to reduce the outliers  
SysBpwinsorize<- Winsorize(Hypertensionn$sysBP, probs = c(0.05,0.95), na.rm = T, type = 1)  
Hypertensionn$sysBP<- SysBpwinsorize

#### Checking for outliers in the diastolic blood pressure variable

Hypertensionn %>%  
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%   
 ggplot(aes(Hypertensionrisk,diaBP))+  
 geom\_boxplot(aes(fill = Hypertensionrisk ),show.legend = F)+  
 labs(title = "Boxplot of Diastolic Blood pressure")

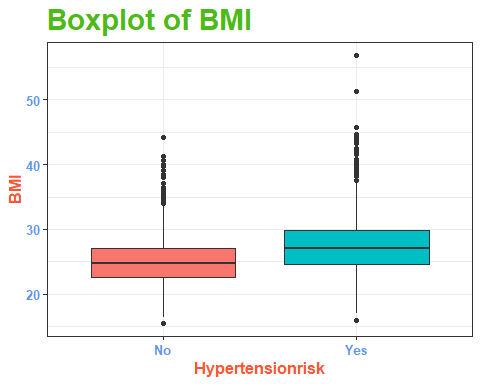


It can be seen from the above plot that Diastolic blood pressure has outliers, we will winsorize the outliers

# winsorizing the diastolic blood pressure in order to reduce the outliers  
DiaBpwinsorize<- Winsorize(Hypertensionn$diaBP, probs = c(0.05,0.95), na.rm = T, type = 1)  
Hypertensionn$diaBP<- DiaBpwinsorize

#### Checking for outliers in the BMI variable

Hypertensionn %>%  
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%   
 ggplot(aes(Hypertensionrisk,BMI))+  
 geom\_boxplot(aes(fill = Hypertensionrisk ),show.legend = F)+  
 labs(title = "Boxplot of BMI")



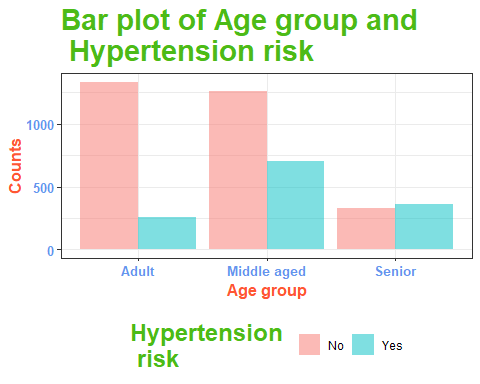
From the above plot the variable BMI contains outliers, we will winsorize it.

# winsorizing the diastolic blood pressure in order to reduce the outliers  
BMIwinsorize<- Winsorize(Hypertensionn$BMI, probs = c(0.05,0.95), na.rm = T, type = 1)  
Hypertensionn$BMI<- BMIwinsorize

## 4. Descriptive statistics

### Age Group Distribution and Hypertension prevalence

Hypertensionn %>%  
 mutate(age\_group = case\_when(  
 age >= 0 & age < 17.9 ~ "Child",  
 age >= 18 & age < 45.9 ~ "Adult",  
 age >= 46 & age < 59.9 ~ "Middle aged",  
 TRUE ~ "Senior"  
 )) %>%   
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%   
   
 ggplot(aes(age\_group))+  
 geom\_bar(aes(fill = Hypertensionrisk), position = "dodge", stat = "count", show.legend = T,alpha = .5)+  
 labs(title = "Bar plot of Age group and \n Hypertension risk",  
 x = "Age group",  
 y = "Counts",  
 fill = "Hypertension \n risk")+  
 theme(legend.position = "bottom")



## Grouping by the age group  
Hypertensionn %>%  
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%  
 group\_by(age\_group) %>%   
 count(Hypertensionrisk) %>% head(5)

# A tibble: 5 × 3  
# Groups: age\_group [3]  
 age\_group Hypertensionrisk n  
 <fct> <fct> <int>  
1 Adult No 1334  
2 Adult Yes 255  
3 Senior No 327  
4 Senior Yes 363  
5 <NA> No 1262

### BMI Distribution

quantile(Hypertensionn$BMI, c(0.1,0.2,0.30,0.4,0.5,0.6,0.7,0.8,0.9,0.95))

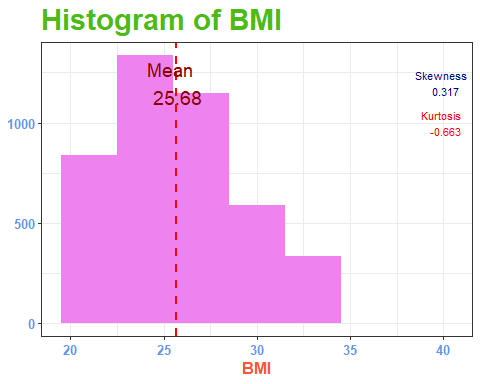
10% 20% 30% 40% 50% 60% 70% 80% 90% 95%   
21.080 22.530 23.570 24.470 25.380 26.350 27.420 28.690 30.761 32.770

shapiro.test(Hypertensionn$BMI) # checking the normality

Shapiro-Wilk normality test  
  
data: Hypertensionn$BMI  
W = 0.96802, p-value < 2.2e-16

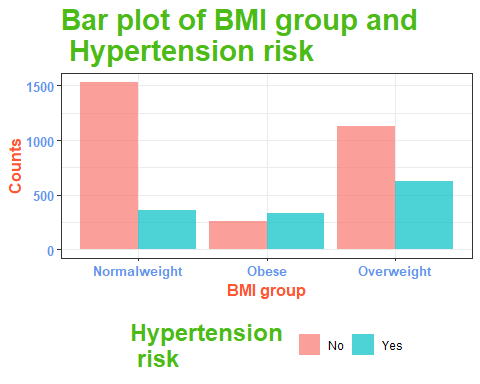
BMI variable is’nt normally distributed because the p-value is less than 0.05.

## BMI Historam  
Hypertensionn %>%   
 ggplot(aes(BMI))+  
 geom\_histogram(binwidth = 3, fill = "#EE82EE")+  
 geom\_vline(aes(xintercept = mean(BMI)), color = "red", size = 1.0, linetype = "dashed")+  
 annotate("text",x = 25.5, y = 1200,  
 label = paste("Mean \n ", round(mean(Hypertensionn$BMI),2)),  
 color = "darkred",  
 size = 5)+  
 annotate("text",x = 40, y = 1200,  
 label = paste("Skewness \n ", round(skewness(Hypertensionn$BMI),3)),  
 color = "navyblue",  
 size = 3)+  
 annotate("text",x = 40, y = 1000,  
 label = paste("Kurtosis \n ", round(kurtosis(Hypertensionn$BMI),3)),  
 color = "red",  
 size = 3)+  
 labs(title = "Histogram of BMI",  
 x = "BMI",  
 y = NULL)



### Bar plot distribution of BMI group and hypertension prevalence

Hypertensionn %>%   
 mutate(bmi\_group = case\_when(  
 BMI < 18.5 ~ "Underweight",  
 BMI >= 18.5 & BMI <= 24.9 ~ "Normalweight",  
 BMI >= 25 & BMI <= 29.9 ~ "Overweight",  
 TRUE ~ "Obese"  
 )) %>%   
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%   
   
 ggplot(aes(bmi\_group))+  
 geom\_bar(aes(fill = Hypertensionrisk), position = "dodge", stat = "count", show.legend = T,alpha = .7)+  
 labs(title = "Bar plot of BMI group and \n Hypertension risk",  
 x = "BMI group",  
 y = "Counts",  
 fill = "Hypertension \n risk")+  
 theme(legend.position = "bottom")



## Grouping by BMI Group  
Hypertensionn %>%  
 mutate(Hypertensionrisk = factor(Hypertensionrisk,  
 levels = c(0 ,1),  
 labels = c("No",  
 "Yes"))) %>%  
 group\_by(bmi\_group) %>%   
 count(Hypertensionrisk) %>% head(5)

# A tibble: 5 × 3  
# Groups: bmi\_group [3]  
 bmi\_group Hypertensionrisk n  
 <fct> <fct> <int>  
1 Obese No 273  
2 Obese Yes 336  
3 Overweight No 1124  
4 Overweight Yes 617  
5 Normalweight No 1475

### Age Distribution

quantile(Hypertensionn$age, c(0.1,0.2,0.30,0.4,0.5,0.6,0.7,0.8,0.9,0.95))

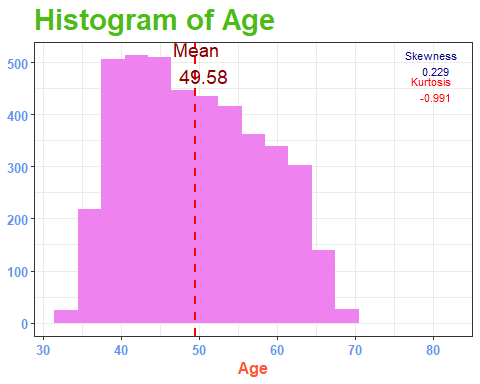
10% 20% 30% 40% 50% 60% 70% 80% 90% 95%   
 39 41 44 46 49 52 55 58 62 64

shapiro.test(Hypertensionn$age) # checking the normality

Shapiro-Wilk normality test  
  
data: Hypertensionn$age  
W = 0.9669, p-value < 2.2e-16

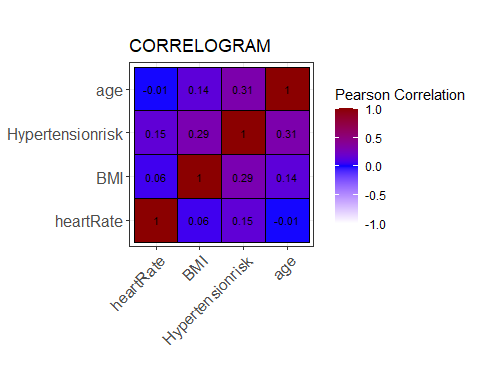
Age is not normally distributed since the p-value is less than 0.05

Hypertensionn %>%   
 ggplot(aes(age))+  
 geom\_histogram(binwidth = 3, fill = "#EE82EE")+  
 geom\_vline(aes(xintercept = mean(age)), color = "red", size = 1.0, linetype = "dashed")+  
 annotate("text",x = 50, y = 500,  
 label = paste("Mean \n ", round(mean(Hypertensionn$age),2)),  
 color = "darkred",  
 size = 5)+  
 annotate("text",x = 80, y = 500,  
 label = paste("Skewness \n ", round(skewness(Hypertensionn$age),3)),  
 color = "navyblue",  
 size = 3)+  
 annotate("text",x = 80, y = 450,  
 label = paste("Kurtosis \n ", round(kurtosis(Hypertensionn$age),3)),  
 color = "red",  
 size = 3)+  
 labs(title = "Histogram of Age",  
 x = "Age",  
 y = NULL)



## 5. Correlation analysis

corr <-Hypertensionn %>%  
 select(heartRate,BMI, Hypertensionrisk,age)   
  
ggcorrplot(cor(corr), title = "CORRELOGRAM",   
 legend.title = "Pearson Correlation" , lab = TRUE,  
 lab\_col = "black",  
 lab\_size = 3, ggtheme = theme\_bw,  
 outline.color = "black",  
 colors = c("white", "blue", "darkred"))



### Correlation analysis in each age group separately with Hypertension and BMI

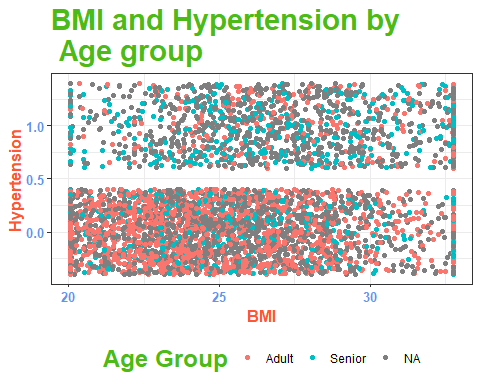
hypert<- Hypertensionn %>%   
 group\_by(age\_group, BMI) %>%   
 summarise(Prevalent = mean(Hypertensionrisk)\*100, .groups = "drop")  
hypertent<- hypert %>%   
 group\_by(age\_group) %>%   
 summarise(correlation = cor(BMI,Prevalent))   
print(hypertent)

# A tibble: 3 × 2  
 age\_group correlation  
 <fct> <dbl>  
1 Adult 0.344  
2 Senior 0.249  
3 Middle 0.305

Their is a weak correlation between the **adult age** group between BMI and Hypertension with **r = 0.344.**

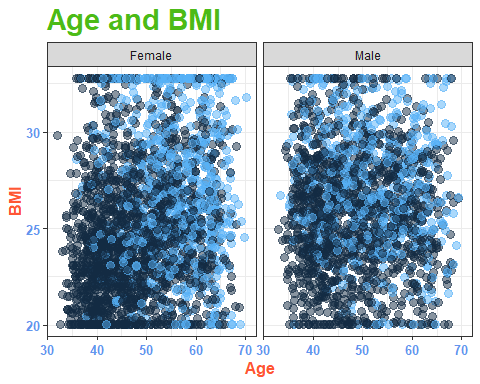
Their is also a weak correlation in the Middle Age Group and the Senior age Group with **r = 0.305 and r = 0.249 respectively.**

Hypertensionn %>%   
 ggplot(aes(x = BMI, y = Hypertensionrisk, color = age\_group))+  
 geom\_jitter()+  
 labs( title = "BMI and Hypertension by \n Age group",  
 x = "BMI",  
 y = "Hypertension ",  
 color = "Age Group")+  
 theme(legend.position = "bottom")



### A scatter plot of Age and BMI faceted by gender

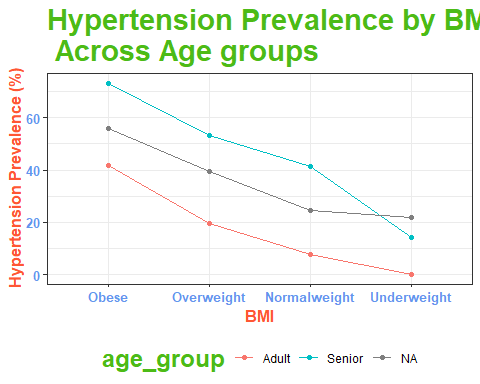
Hypertensionn %>%  
 mutate(gender = factor(gender,  
 levels = c(0 ,1),  
 labels = c("Female",  
 "Male"))) %>%  
 ggplot(aes(x = age,   
 y = BMI,  
 color = Hypertensionrisk))+  
 facet\_wrap(~gender)+  
 geom\_jitter(size = 3,  
 alpha = .5,show.legend = F)+  
 labs(title = "Age and BMI ",  
 x= "Age",  
 y = "BMI")



## 6. Data Visualization

### A line graph to show how Hypertension prevalence increases with BMI across different Age Groups

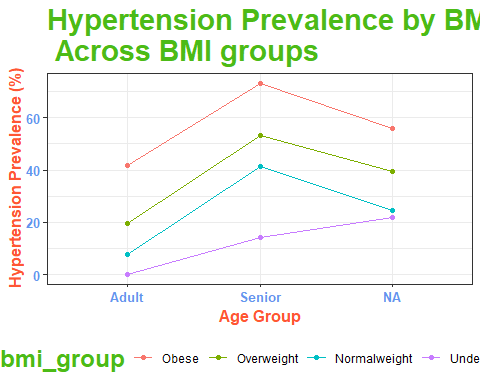
hypertension\_prevalence <- Hypertensionn %>%   
 group\_by(age\_group, bmi\_group) %>%   
 summarise(Prevalence = mean(Hypertensionrisk)\* 100, .groups = "drop")  
  
hypertension\_prevalence %>%   
 ggplot(aes(bmi\_group, Prevalence, group = age\_group, color = age\_group))+  
 geom\_line()+  
 geom\_point()+  
 labs(title = "Hypertension Prevalence by BMI \n Across Age groups",  
 x = "BMI",  
 y = "Hypertension Prevalence (%)")+  
 theme(legend.position = "bottom")



From the graph it can be evidently seen that Hypertension prevalence is high in Obese and the Senior age group (60 Years and above)

### A line graph to show how Hypertension prevalence increases with BMI across different BMI Groups

hypertension\_prevalent <- Hypertensionn %>%   
 group\_by(age\_group, bmi\_group) %>%   
 summarise(Prevalenc = mean(Hypertensionrisk)\* 100, .groups = "drop")  
  
hypertension\_prevalent %>%   
 ggplot(aes(age\_group, Prevalenc, group = bmi\_group, color = bmi\_group))+  
 geom\_line()+  
 geom\_point()+  
 labs(title = "Hypertension Prevalence by BMI \n Across BMI groups",  
 x = "Age Group",  
 y = "Hypertension Prevalence (%)")+  
 theme(legend.position = "bottom")



Also from the graph its evident that Obese individuals have high Hypertention prevalence

## 7. Conclusion

In this study, their is a varying relationship between BMI and Hypertension prevalence across age groups. In the adult age group **(18-45.9 years)**, BMI had a weak correlation with Hypertension. The middle aged**(46-59.9 years)** and senior age group**(above 60 years)** similarly had a weak correlation.

From this Data set BMI seems to be a less reliable indicator in Older Adults.

From the analysis their seems to be other dominant risk factors that influence Hypertension Prevalence, other than BMI.

A low correlation between BMI and Hypertension Prevalence inn seniors could result from multiple factors, such as their influence of other health conditions, the less reliable nature of BMI in older adults, or the presence of other risk factors like: Heart disease and smoking.

However, a line graph shows that Hypertension Prevalence is **higher** in the Obese individuals, followed by Overweight, Normal weight and Under weight respectively.

The line Graph also shows that The Hypertension Prevalence is **Higher** in senior individuals(greater than 60 years), followed by Middle aged(46-59.9 years), Adults(18-45.9 years) and then Children respectively.(0-17.9 years)

## Recommendations

Given the low correlation between BMI and hypertension in seniors, its important to recognize that BMI alone isn’t a sufficient predictor of Hypertension in this Population. Therefore the focus should shift to other factors that may contribute to Hypertension in older adults, for example: Heart disease, Diabetes and smoking.

### Action:

1. A further research should be carried out in order to understand in in-depth these other risk factors.
2. Promoting education on Hypertension risk factors and implement target intervention among the senior age group.