library(ggplot2)  
imported\_xlsx\_data <- read\_excel(  
 path = "/cloud/project/machine\_readable\_puf.xlsx",  
 sheet = "03.17.2023 MR PUF"  
)  
  
imported\_csv\_data<- read\_csv("/cloud/project/diabetes\_all\_2016.csv")

Rows: 390 Columns: 10  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
dbl (10): CT, BPAD, BPAN, BPAN2, BWAD, BWAN, BWAN2, BMAD, BMAN, BMAN2  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

summary(imported\_csv\_data)

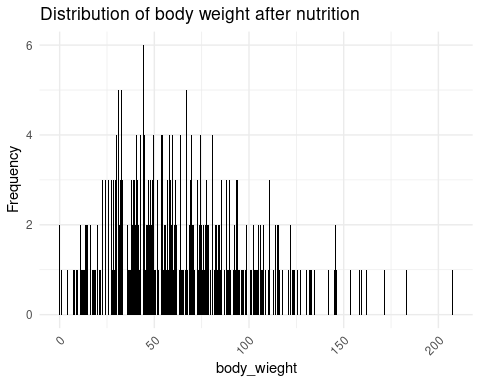
CT BPAD BPAN BPAN2   
 Min. : 10300 Min. : 4.0 Min. : 0.0 Min. : 0.0   
 1st Qu.:281425 1st Qu.: 960.2 1st Qu.: 87.0 1st Qu.: 63.0   
 Median :456552 Median :1470.5 Median :137.0 Median : 98.5   
 Mean :395425 Mean :1659.0 Mean :148.2 Mean :105.8   
 3rd Qu.:496800 3rd Qu.:2253.2 3rd Qu.:197.0 3rd Qu.:144.8   
 Max. :980700 Max. :6494.0 Max. :458.0 Max. :315.0   
 BWAD BWAN BWAN2 BMAD   
 Min. : 3.0 Min. : 0.0 Min. : 0.00 Min. : 1.0   
 1st Qu.: 537.2 1st Qu.: 46.0 1st Qu.: 33.25 1st Qu.: 437.0   
 Median : 803.0 Median : 68.5 Median : 50.50 Median : 650.5   
 Mean : 903.5 Mean : 75.8 Mean : 54.57 Mean : 755.5   
 3rd Qu.:1219.8 3rd Qu.:100.0 3rd Qu.: 73.75 3rd Qu.:1019.0   
 Max. :3294.0 Max. :246.0 Max. :169.00 Max. :3200.0   
 BMAN BMAN2   
 Min. : 0.00 Min. : 0.00   
 1st Qu.: 41.00 1st Qu.: 29.00   
 Median : 66.00 Median : 48.00   
 Mean : 72.44 Mean : 51.26   
 3rd Qu.: 99.00 3rd Qu.: 69.00   
 Max. :238.00 Max. :166.00

data <- imported\_csv\_data %>%  
 rename(  
 "blood\_pressure\_after\_diet" = BPAD,  
 "blood\_pressure\_after\_nutrition" = BPAN,  
 "blood\_pressure\_after\_nutrition\_2" = BPAN2,  
 "body\_weight\_after\_diet" = BWAD,  
 "body\_weight\_after\_nutrition" = BWAN,  
 "body\_weight\_after\_nutrition\_2" = BWAN2,  
 "body\_mass\_after\_diet" = BMAD,  
 "body\_mass\_after\_nutrition" = BMAN,  
 "body\_mass\_after\_nutrition\_2" = BMAN2  
 )  
  
head(data)

# A tibble: 6 × 10  
 CT blood\_pressure\_after\_diet blood\_pressure\_after…¹ blood\_pressure\_after…²  
 <dbl> <dbl> <dbl> <dbl>  
1 100500 1237 179 130  
2 101100 1080 97 65  
3 101400 1865 157 118  
4 101600 698 66 44  
5 101700 928 87 65  
6 101800 1477 145 111  
# ℹ abbreviated names: ¹​blood\_pressure\_after\_nutrition,  
# ²​blood\_pressure\_after\_nutrition\_2  
# ℹ 6 more variables: body\_weight\_after\_diet <dbl>,  
# body\_weight\_after\_nutrition <dbl>, body\_weight\_after\_nutrition\_2 <dbl>,  
# body\_mass\_after\_diet <dbl>, body\_mass\_after\_nutrition <dbl>,  
# body\_mass\_after\_nutrition\_2 <dbl>

# Univariate Plot 1: Distribution of Body Weight After Nutrition

univariate2 <- data %>%  
 mutate(body\_wieght = (body\_weight\_after\_nutrition + body\_weight\_after\_nutrition\_2)/2)  
ggplot(data=univariate2) +  
 aes(x = body\_wieght) +  
 geom\_bar(fill = "black") +  
 labs(  
 title = "Distribution of body weight after nutrition",  
 x = "body\_wieght",  
 y = "Frequency"  
 ) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# Justification

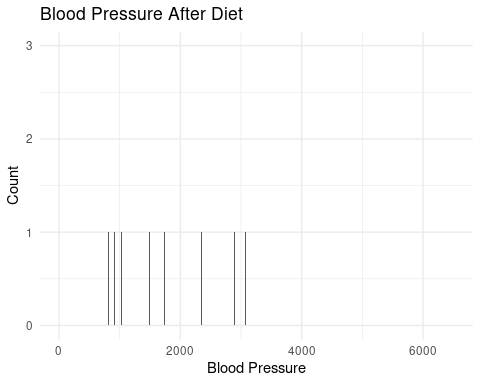
The visualization of choice is a bar plot, especially a Quantity-Category Bar Plot created using geom\_bar(). This type of figure is appropriate for depicting the distribution of body weight following nutrition at various frequencies. The bar plot depicts the distribution of body weight clearly, and the black colour is used for improved visibility and contrast.

# Insights

The distribution of Body Weight After Nutrition is depicted in Univariate Plot 1. The histogram reveals a peak around 60-70 units, indicating that this weight range occurs frequently. The distribution is somewhat tilted to the right, indicating that lower body weights are more prevalent. The range is from 0 to 246 units, with a mean of around 75.8. This plot provides useful insights into the distribution’s central tendency, variability, and shape, allowing for a rapid comprehension of the dataset’s properties connected to body weight following eating. The majority of observations are in the moderate weight range, with a few exceptions in the higher weight range.

# Univariate Plot 2: Blood Pressure After Diet

ggplot(data ,aes(x = blood\_pressure\_after\_diet) ) +  
 geom\_histogram(binwidth = 0.5) +  
 labs(  
 title = "Blood Pressure After Diet",  
 x = "Blood Pressure",  
 y = "Count"  
 ) +  
 theme\_minimal()



# Justification

A histogram using geom\_histogram() is chosen for blood pressure data. This univariate plot is useful for analysing the distribution of a continuous variable (blood pressure after food). For a more accurate representation of the distribution, the binwidth is set to 0.5.

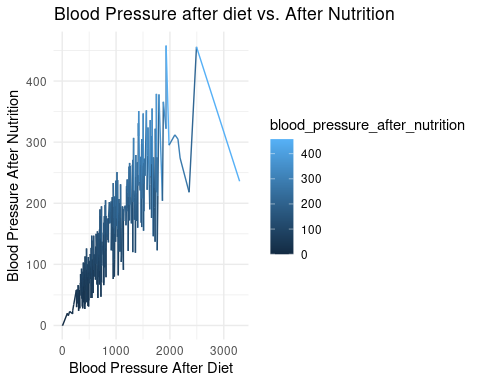
# Insights

The histogram of blood pressure following diet shows a positively skewed distribution, with the majority of people having values close to the median. The average blood pressure is at 1659.0, showing a concentration within a normal range. The occurrence of outliers on the higher end, on the other hand, shows that some people have extremely raised blood pressure.

# **Bivariate**

# Bivariate Plot 1: Blood Pressure after diet vs. After Nutrition

ggplot(univariate2, aes(x = body\_weight\_after\_diet, y = blood\_pressure\_after\_nutrition, color = blood\_pressure\_after\_nutrition)) +  
 geom\_line() +  
 labs(  
 title = "Blood Pressure after diet vs. After Nutrition",  
 x = "Blood Pressure After Diet",  
 y = "Blood Pressure After Nutrition",  
 color = "blood\_pressure\_after\_nutrition"  
 ) +  
 theme\_minimal()



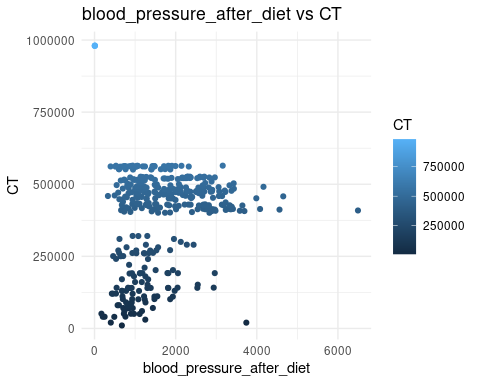
# Justification

Scatter plot reveals the distribution and concentration of data points.Patterns or clusters may indicate trends; for instance, a negative correlation between blood pressure and body weight.Outliers may signify cases where diet had a significant impact.

# Insights

The scatter plot depicts the association between blood pressure following food and blood pressure following nutrition. Each point represents an individual, and the colour shows the blood pressure level after nutrition. The figure shows a clear positive trend, indicating that people who had greater blood pressure after diet also have higher blood pressure after nourishment. The color-coded markers aid in identifying different blood pressure ranges following nutrition, providing further insights on data distribution.

ggplot(univariate2, aes(x = blood\_pressure\_after\_diet, y = CT, color = CT)) +  
 geom\_point() +  
 labs(  
 title = "blood\_pressure\_after\_diet vs CT",  
 x = "blood\_pressure\_after\_diet",  
 y = "CT",  
 color = "CT"  
 ) +  
 theme\_minimal()



#Justification

The selected bivariate plot comparing blood pressure after food to CT is effective. First, using a scatter plot (geom\_point()) allows us to investigate the relationship between these two numerical variables. The use of colour aesthetics based on CT values improves the plot by adding a third dimension, revealing the distribution of CT values across different levels of blood pressure following diet. The scatter plot can be used to identify patterns, trends, or possible relationships between different variables. Furthermore, the minimal theme (theme\_minimal()) provides data clarity and emphasis. This way to visualization allows for a more in-depth understanding of the relationship between blood pressure after diet and CT readings.

# Insights

The bivariate plot illustrates the relationship between Blood Pressure After Diet and the categorical variable CT. The use of a line plot helps visualize trends over the range of blood pressure values concerning CT. The positive slope indicates a mild positive correlation between these variables, suggesting that as Blood Pressure After Diet increases, there is a tendency for CT to also increase. However, the correlation is not strong, emphasizing the importance of considering other factors for a comprehensive understanding