

# Nature or Nurture?


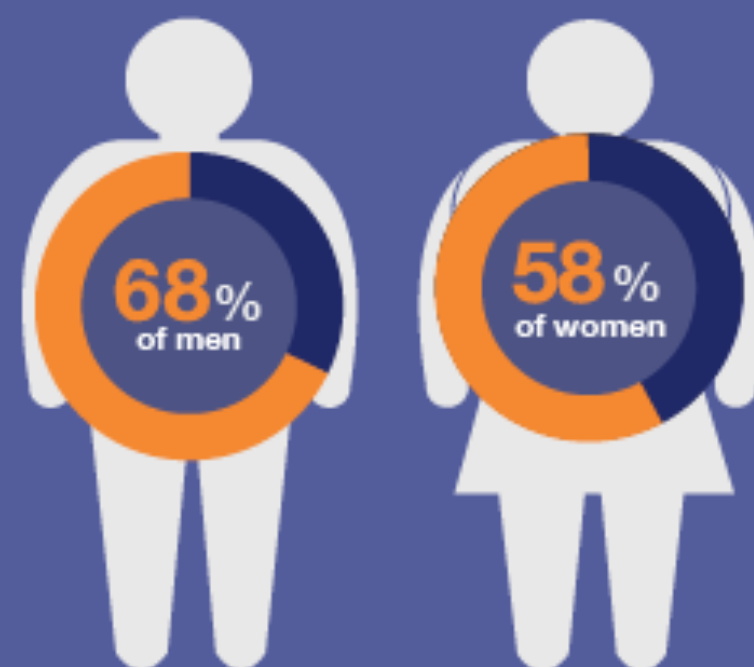
Estimating Obesity Based on Physical Condition & Eating Habits

---

**STAT3622 Final Project Presentation by  
Group 6**

Soo-ah Kim (3035661061) &  
Dongjun Yeom (3035666463)

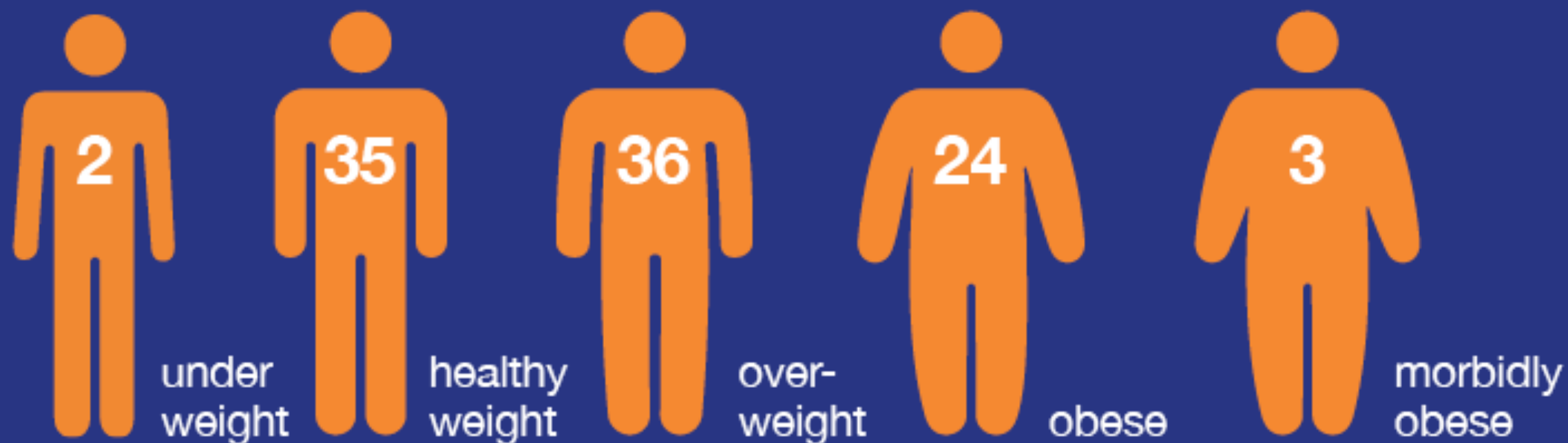
In 2015  
**63%**  
of adults in England  
were **overweight**  
or **obese**



In England, the prevalence of obesity  
among adults rose from **14.9%** to  
**26.9%** between 1993 and 2015

The gauge shows a needle pointing from 14.9% to 26.9%.

Of every  
**100 adults**  
in England  
there are...



# Estimation of Obesity Levels Based On Eating Habits and Physical Condition

## Physical Condition

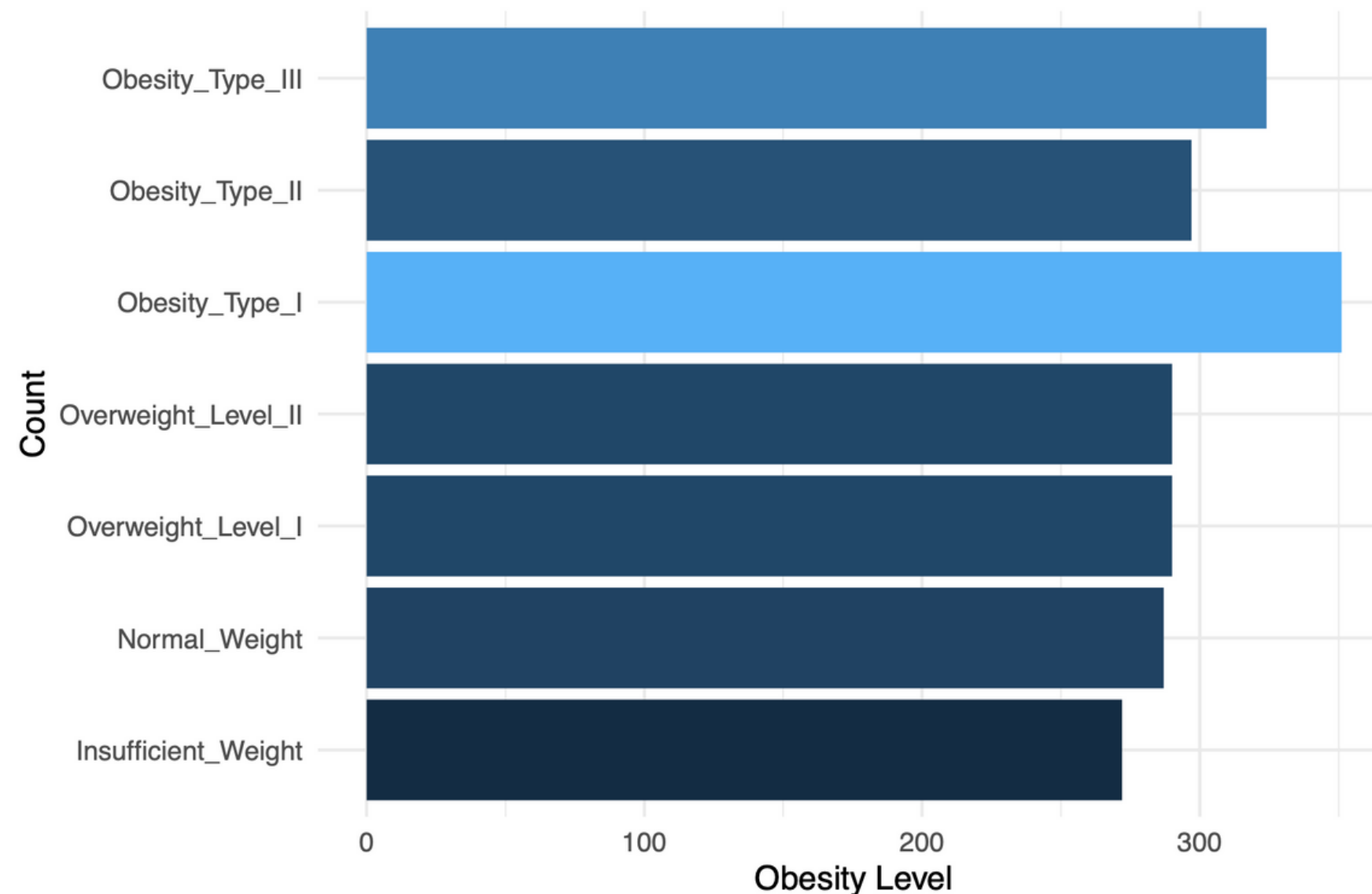
- Gender
- Age
- Height
- Family History with Obesity
- Body Mass Index (BMI)

## Habits

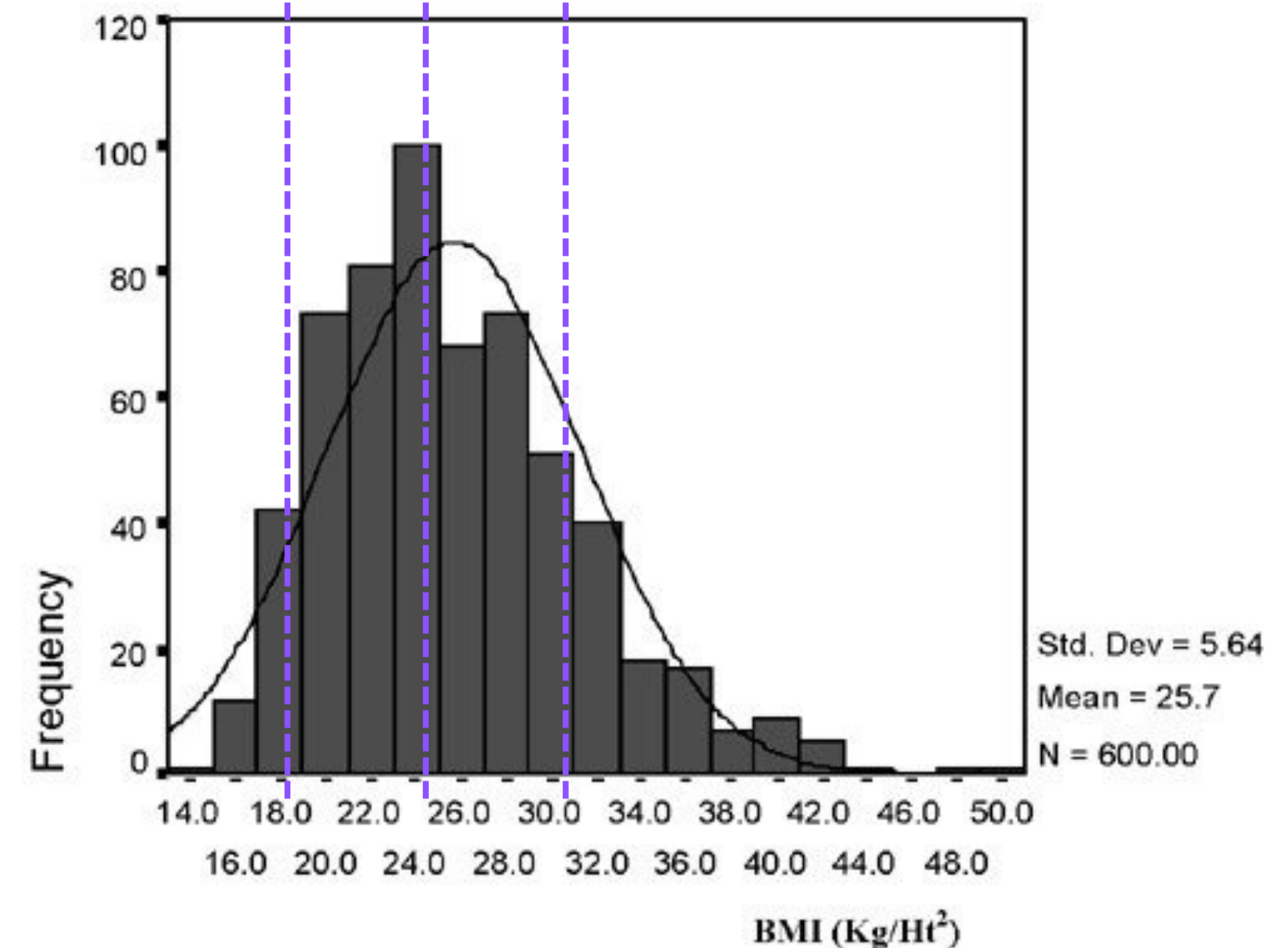
- Frequent consumption of high caloric food (FAVC)
- Frequency of consumption of vegetables (FCVC)
- Number of main meals (NCP)
- Consumption of food between meals (CAEC)
- Consumption of water daily (CH20)
- Consumption of alcohol (CALC)
- Calories consumption monitoring (SCC)
- Physical activity frequency (FAF)
- Time using technology devices (TUE)
- Transportation used (MTRANS)

# Distribution of Obesity Levels

## Distribution of Obesity Levels



## Al-Malki et al., 2003

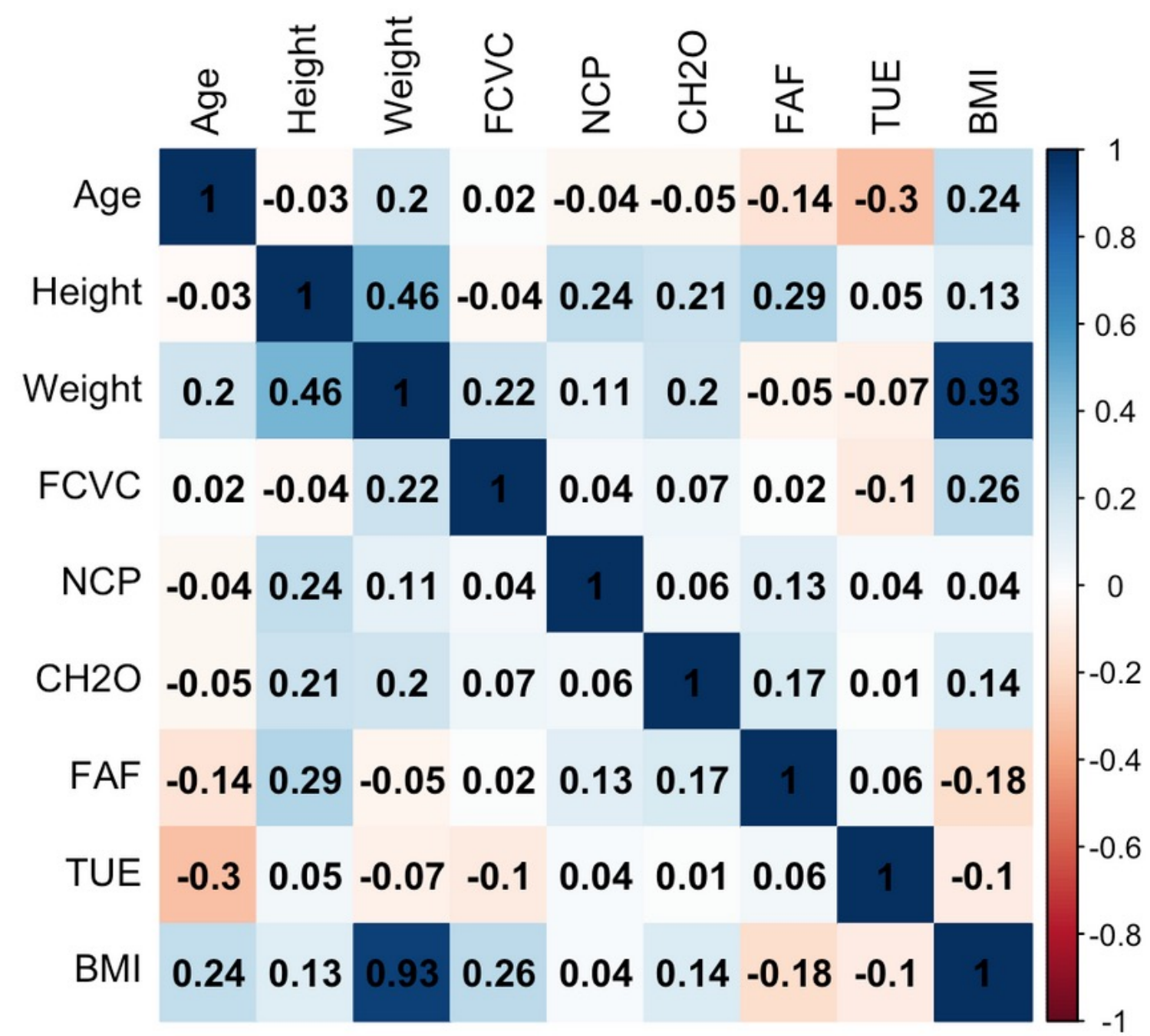


- No class imbalance
- Might indicate a sampling bias due to difference with the real world distribution

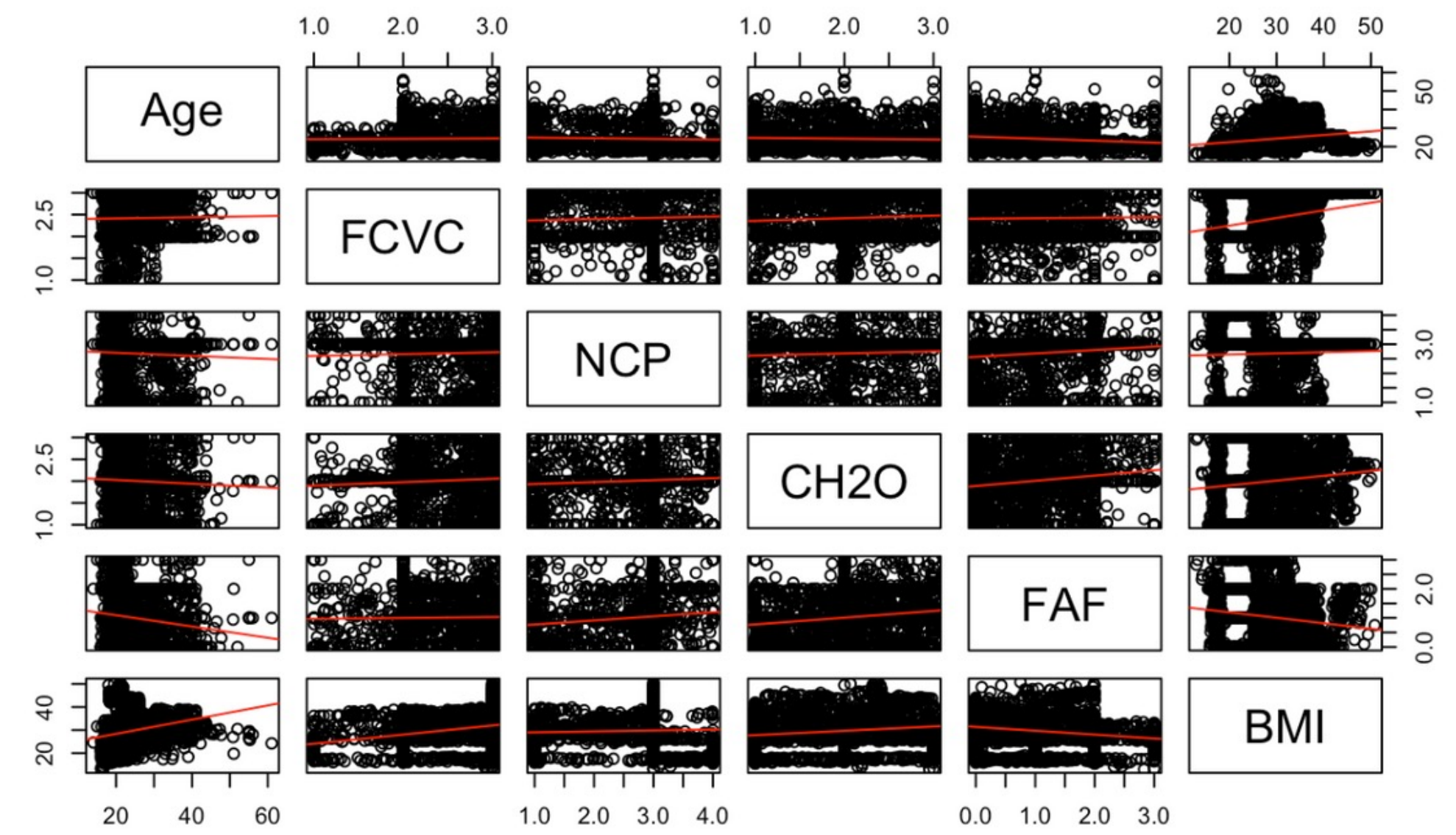


# Correlation between variables

Correlation Plot



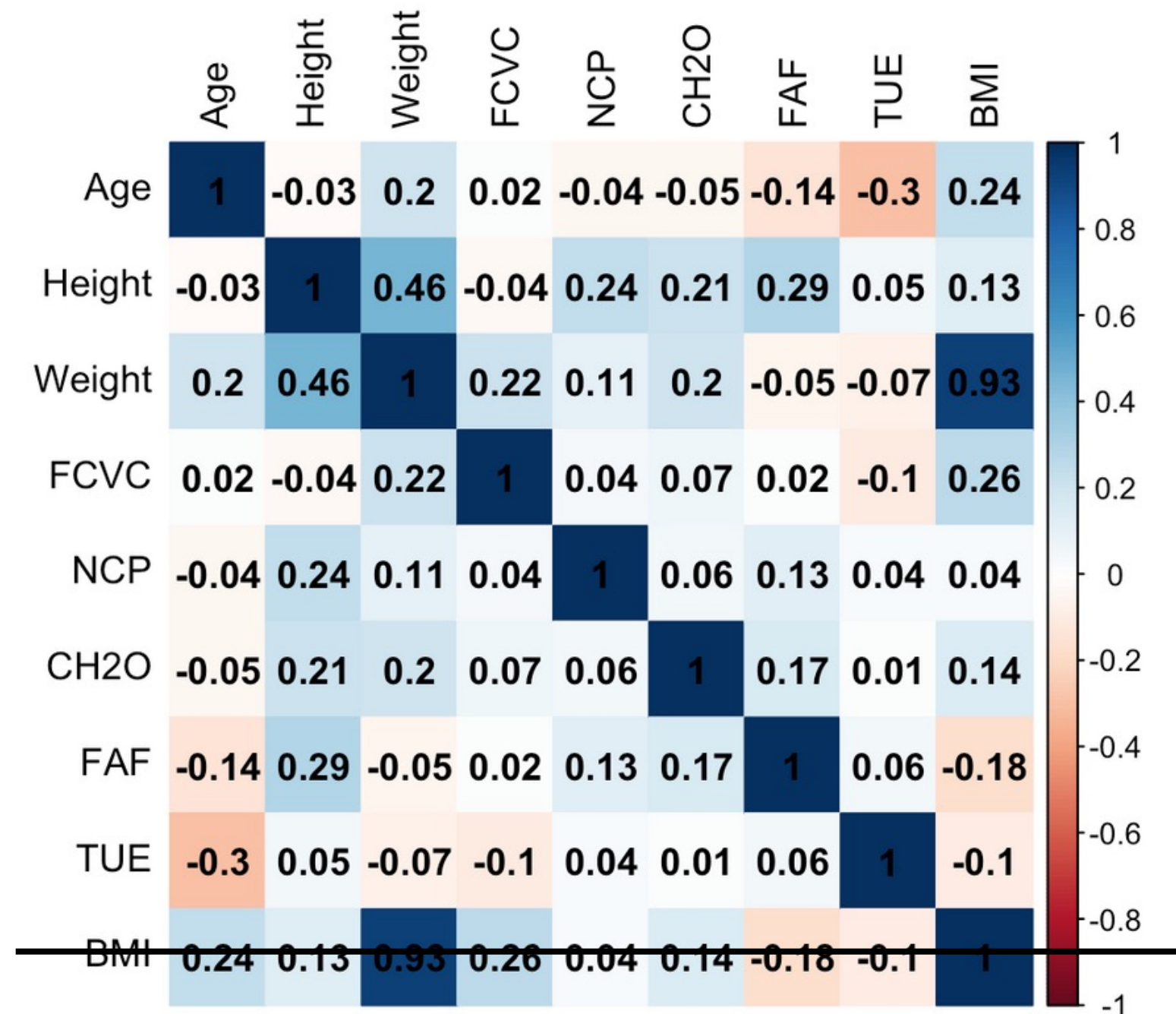
Pairwise Plot



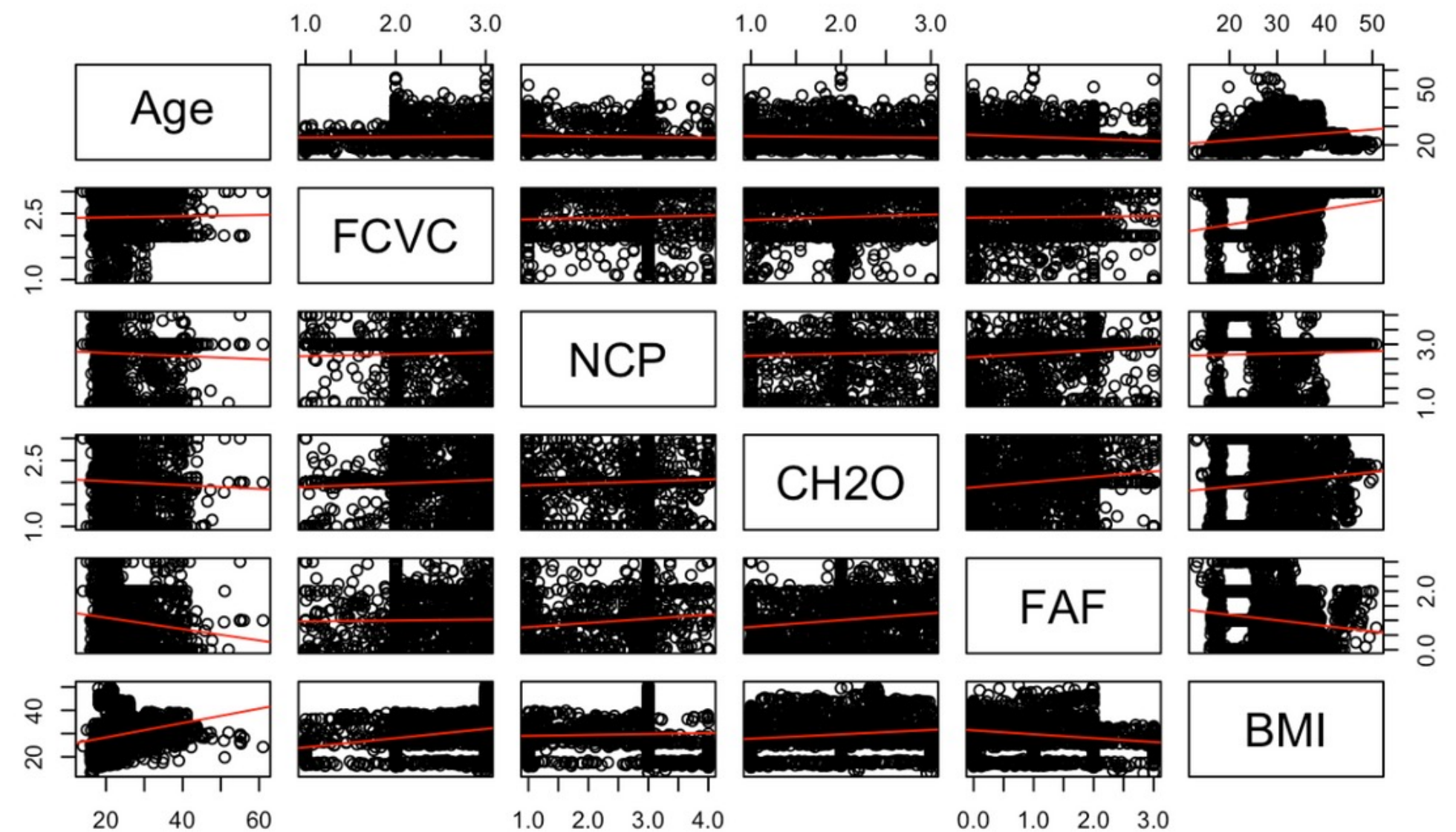


# Correlation between variables

Correlation Plot



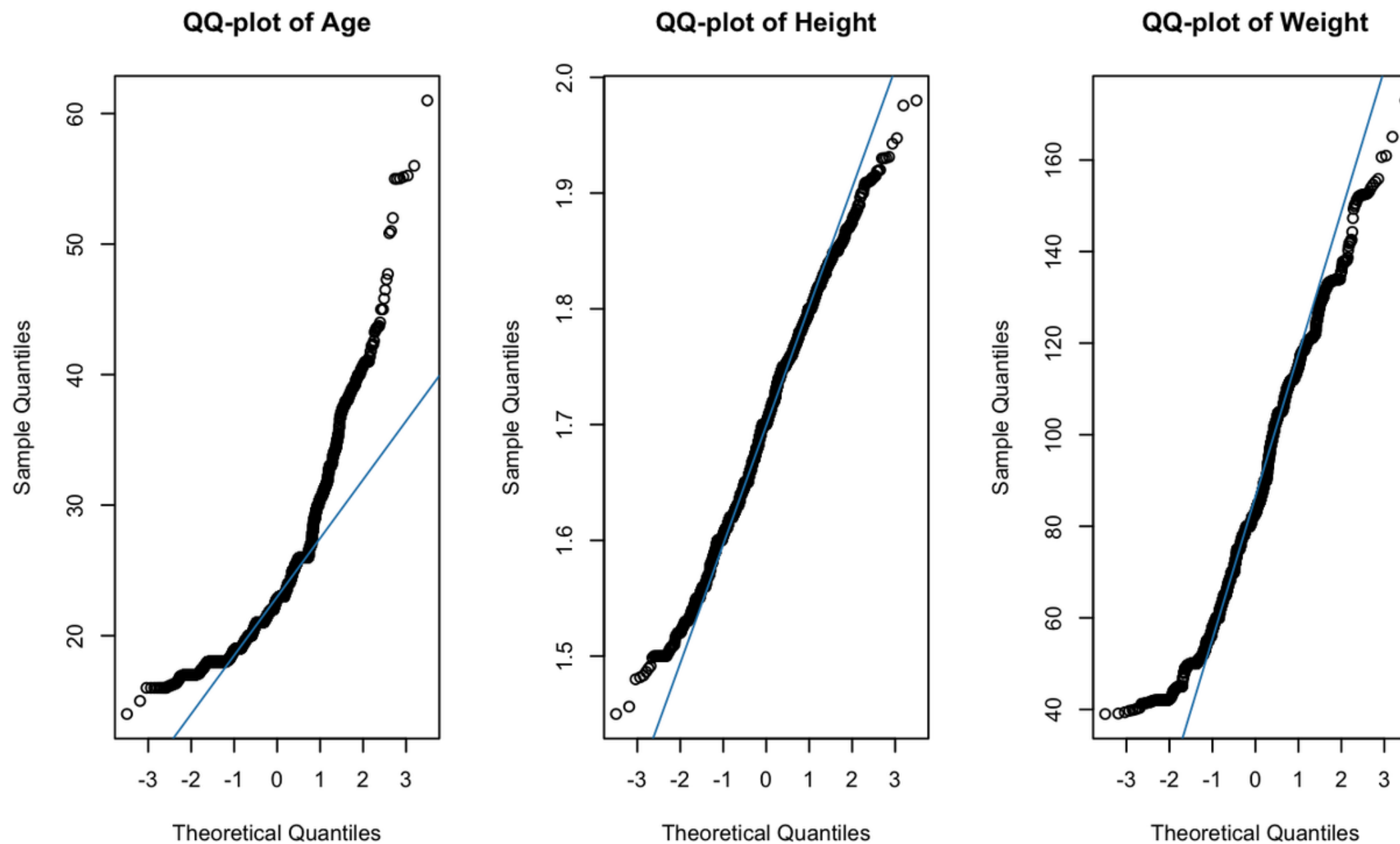
Pairwise Plot



- due to multicollinearity

# Distribution of numeric variables

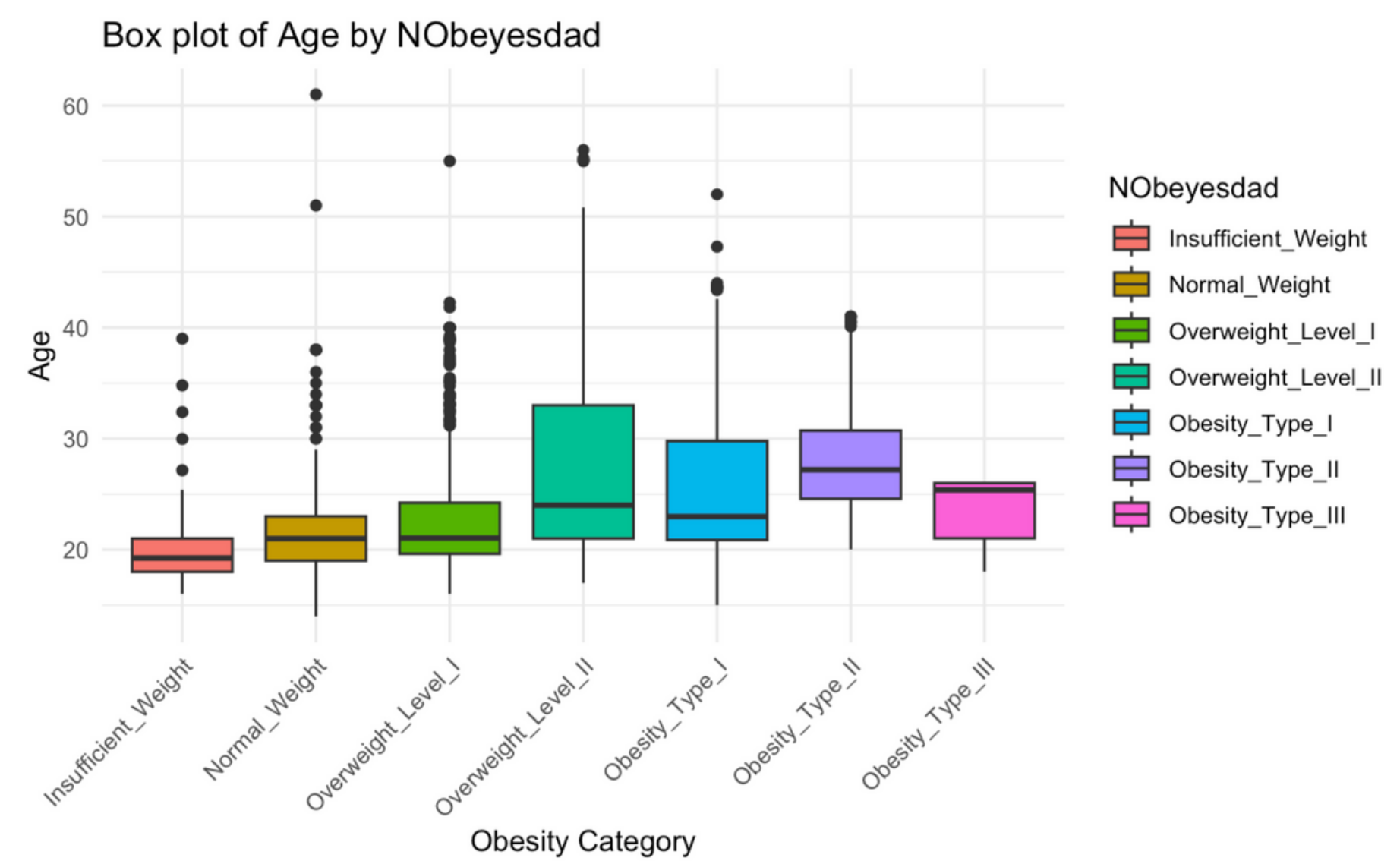
## QQ-Plots



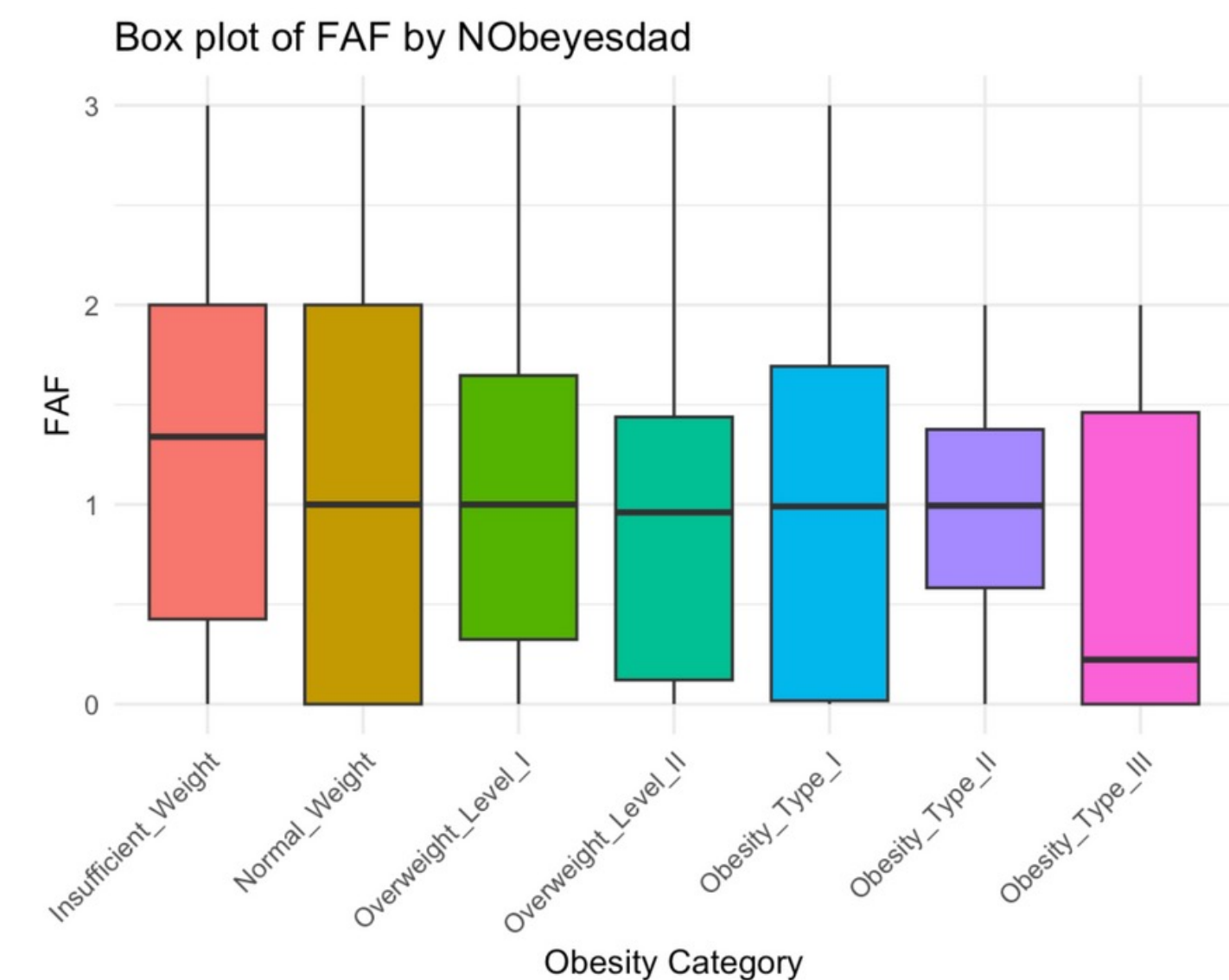
- **Height** and **Weight** roughly follow the normal distribution
- **Age** does not
- **Shapiro-Wilk Normality Test**
  - $W = 0.86606$
  - $p\text{-value} < 2.2e-16$
  - suggests a noticeable deviation from the normal distribution
- **Kurtosis**
  - 5.816858
  - distribution has heavier tails than a normal distribution ( $=3$ )
- **Solution:** avoid models that assume normality or normalize using log or sqrt transformations

# Box Plots by Obesity level

## Age



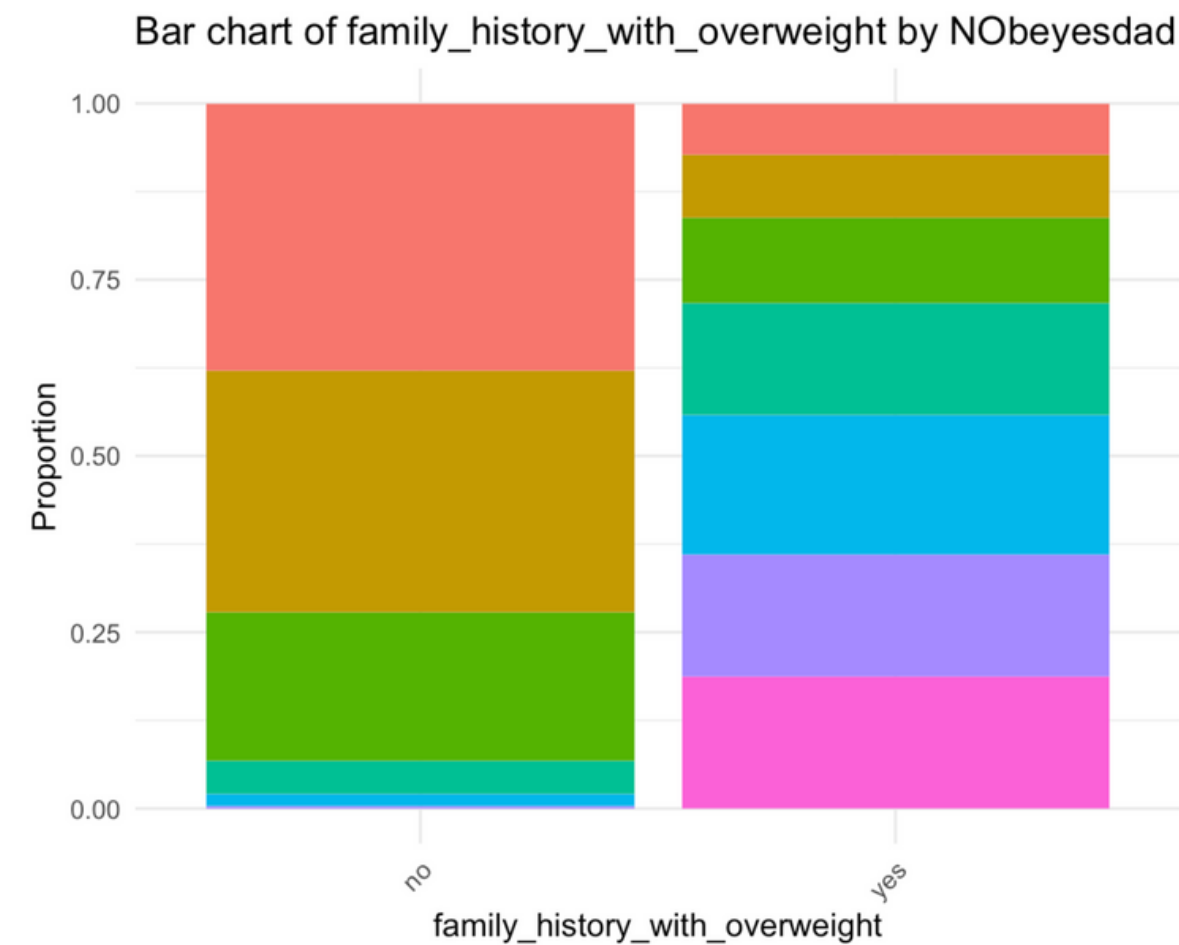
## FAF (Frequency of Physical Activity)



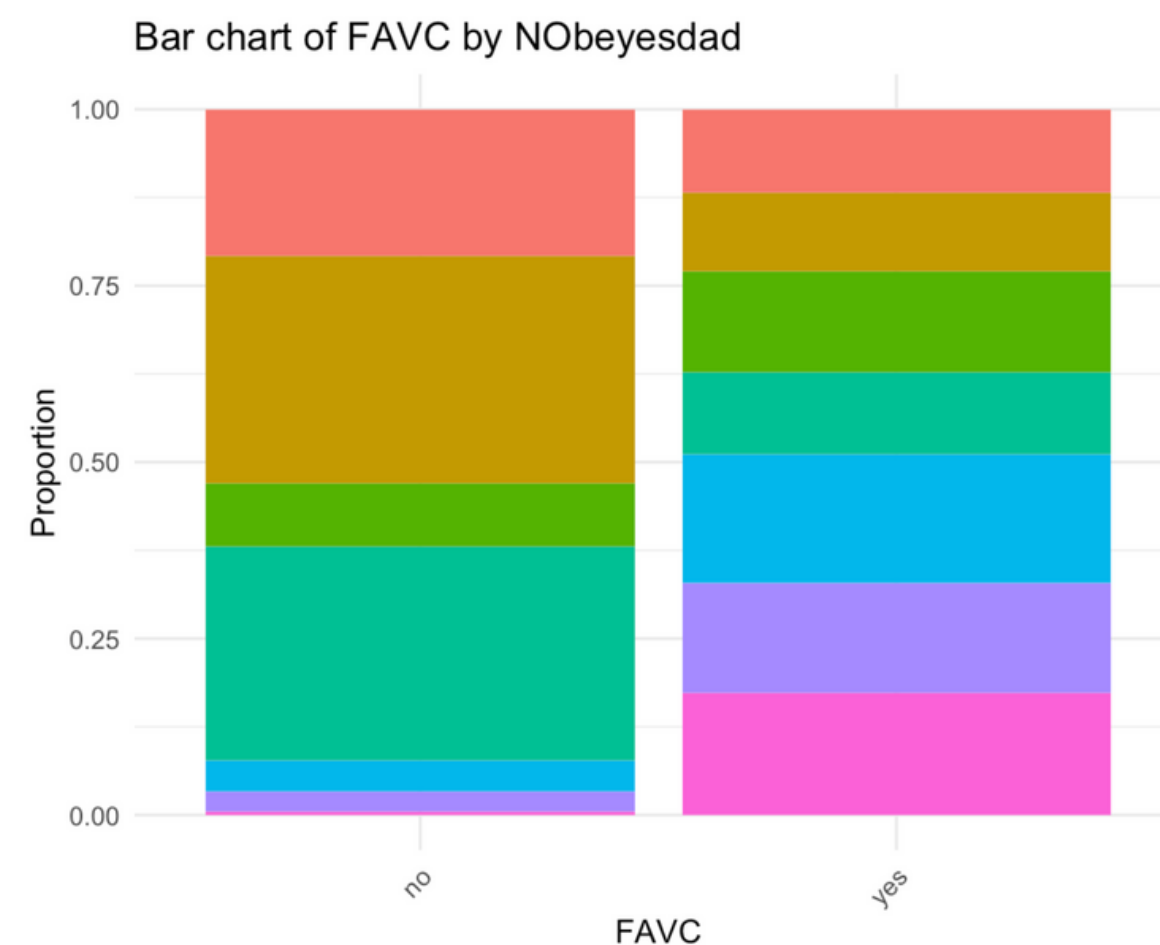


# Bar Charts by Obesity level

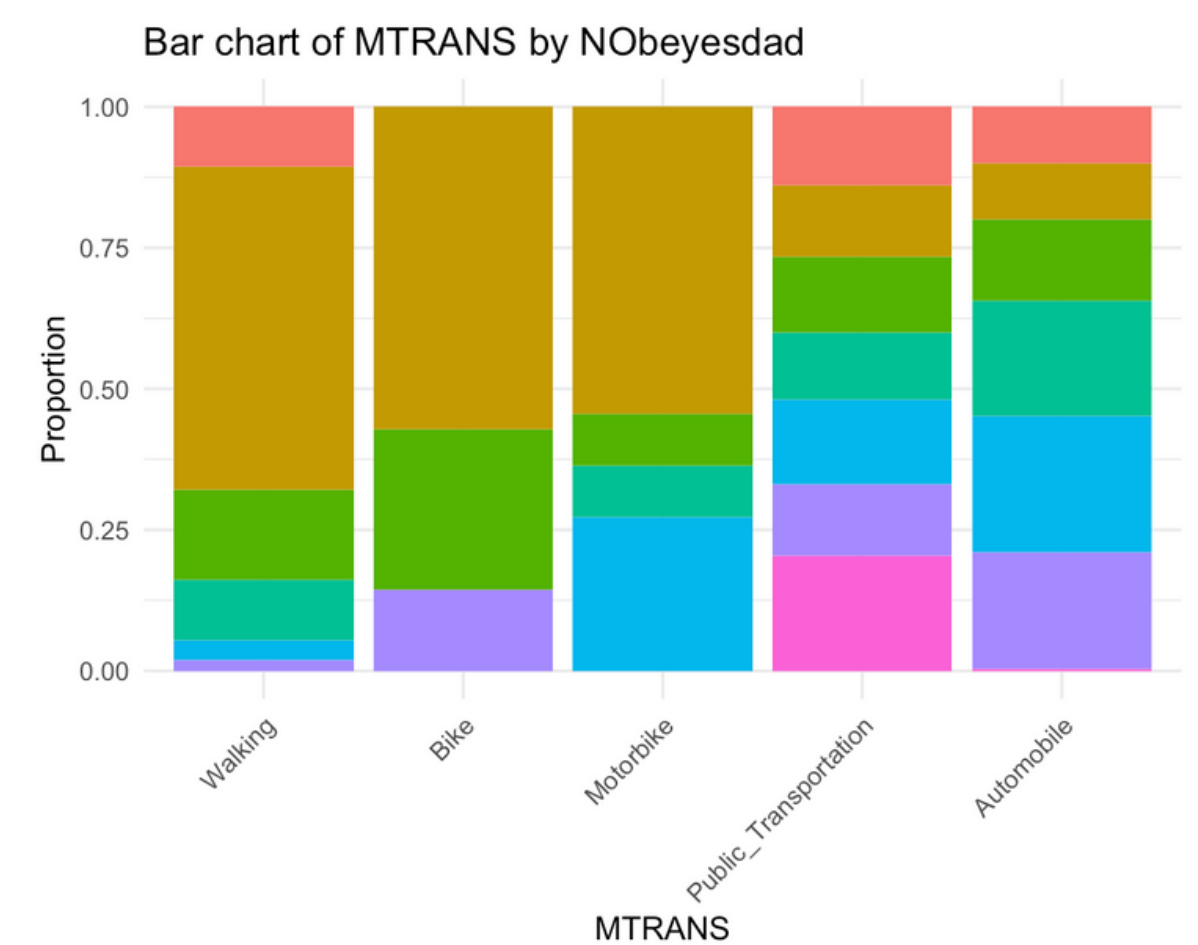
## Family history of obesity



## Frequent consumption of high Caloric food



## Mode of transportation



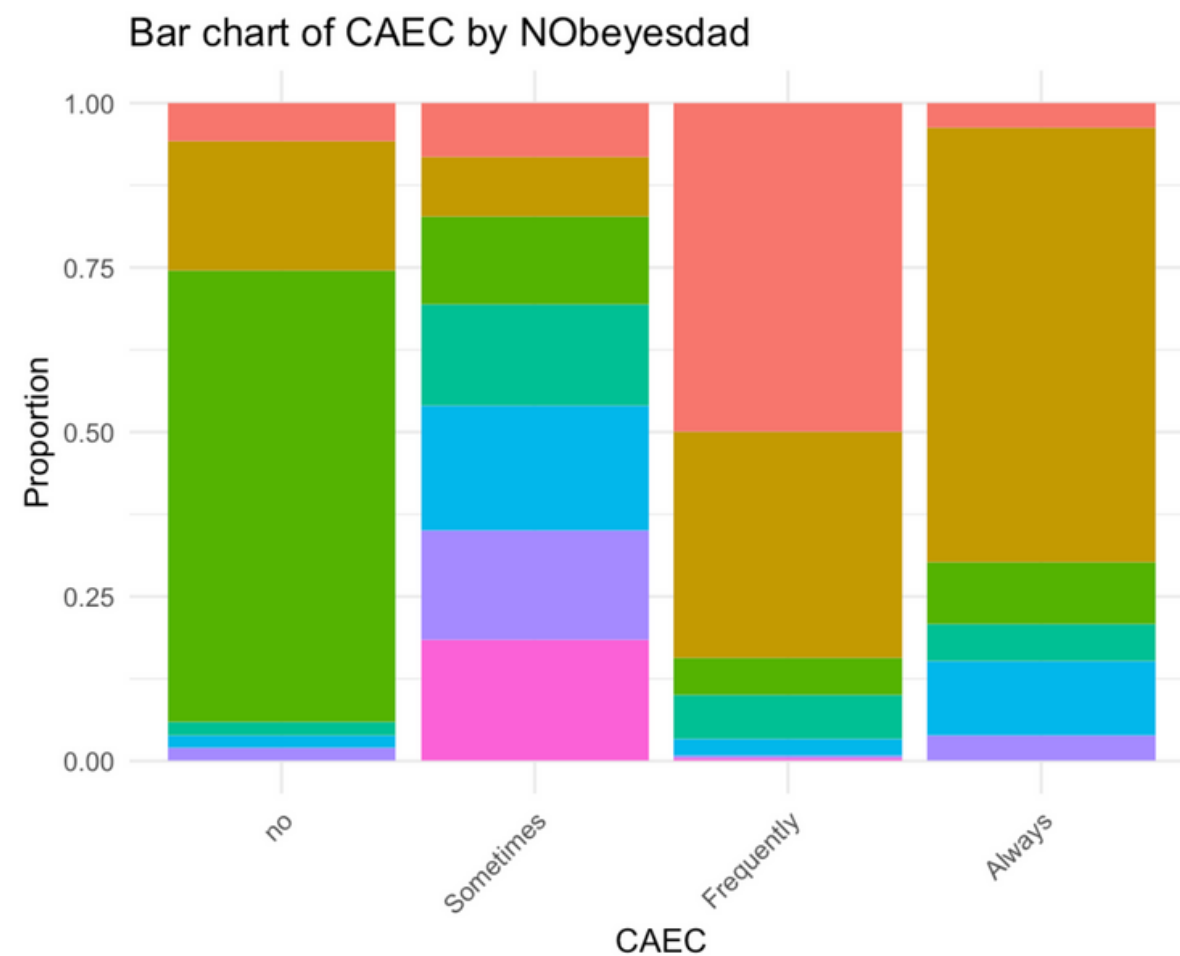
## Interesting bar charts

- People with obesity tend to have **family history** of obesity
- People with obesity tend to **frequently consume high caloric food**
- **Mode of transportation** is also highly related to obesity!

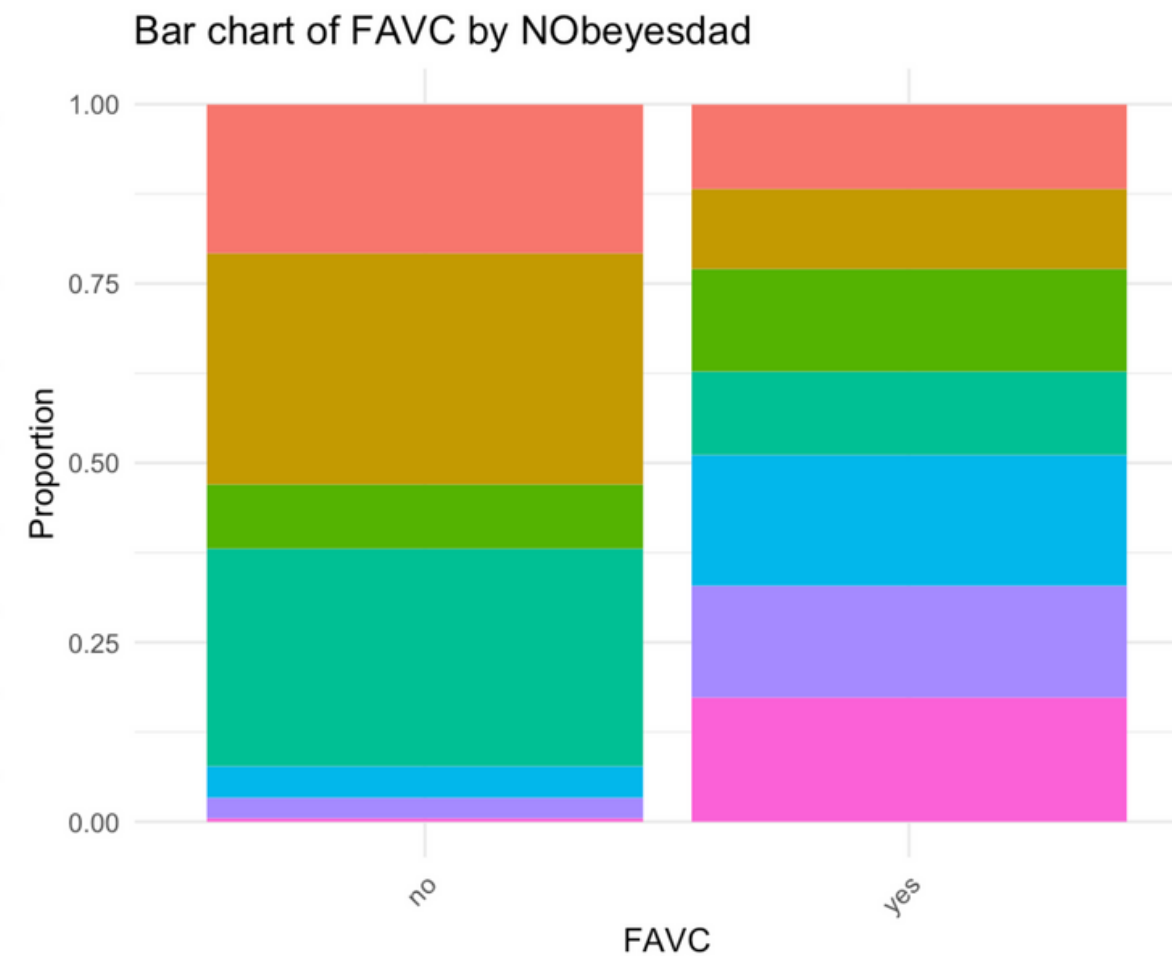


# Bar Charts by Obesity level

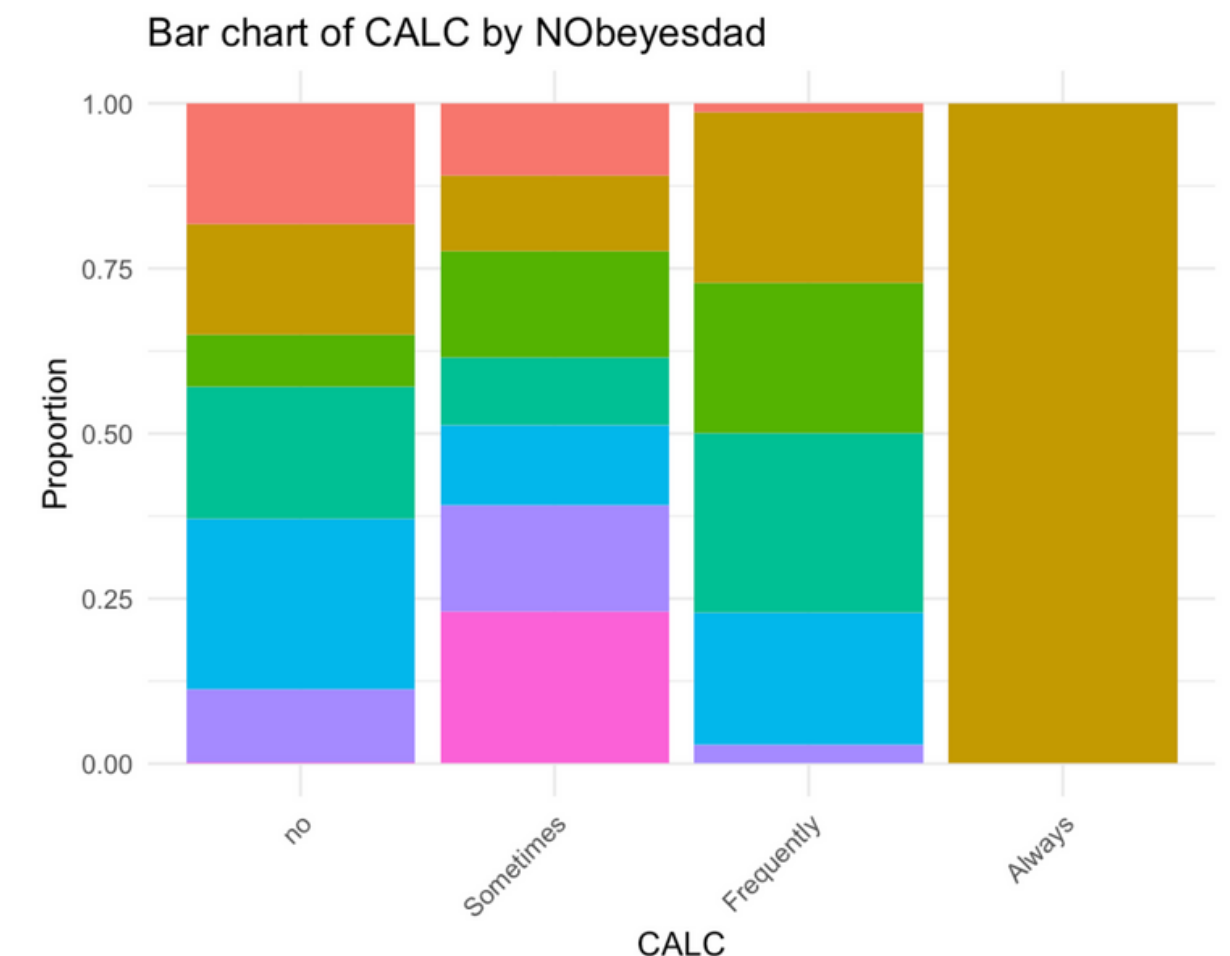
## Consumption of food between meals



## Frequent consumption of vegetables



## Consumption of Alcohol



### Interesting bar charts (counterintuitive?)

- Overweight or obese people tend to not or only sometimes **consume food between meals** while underweight or normal weight people tend to frequently or always eat between meals
- People with obesity tend to more **frequently consume vegetables**
- Overweight or obese people tend to not or only sometimes **consume alcohol**



# Summary of EDA

---

## Removed columns due to multicollinearity

- BMI: too highly related to weight & height

## Are the numeric variables normally distributed?

- Age is not normally distributed
- avoid models that assume normality



# Multinomial Logistic Regression

## Consumption of Food between Meals

|                     | NCP      | CAEC     | Sometimes | CAEC         | Frequently | CAEC | Always |
|---------------------|----------|----------|-----------|--------------|------------|------|--------|
| Normal_Weight       | 1.130728 | 3.752709 | 3.703433  | 5.286345e+00 |            |      |        |
| Overweight_Level_I  | 1.219140 | 3.512026 | 3.287176  | 4.296967e+00 |            |      |        |
| Overweight_Level_II | 1.313788 | 3.539225 | 3.340290  | 3.464077e+00 |            |      |        |
| Obesity_Type_I      | 1.537958 | 5.153274 | 5.397400  | 4.747190e+00 |            |      |        |
| Obesity_Type_II     | 4.168705 | 5.814312 | 5.142472  | 4.460149e+00 |            |      |        |
| Obesity_Type_III    | 3.002694 | 3.365435 | 5.639733  | 9.653115e-39 |            |      |        |

Residual Deviance: 206.2212  
AIC: 506.2212

Degrees of freedom  
= (Number of Observations) – (Number of Parameters)  
= 2111 - 24 = 2087

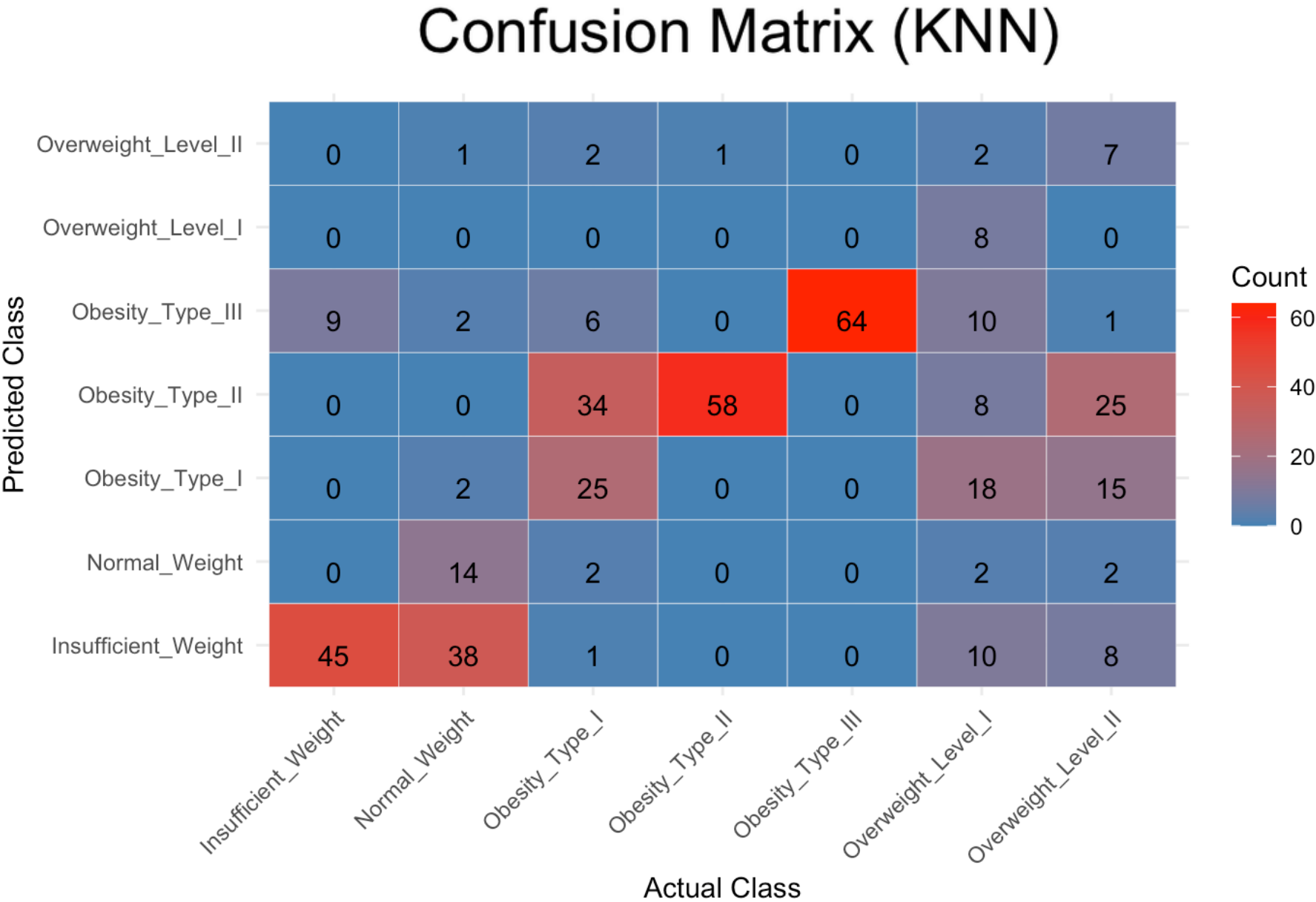
## Consumption of Alcohol

|                     | CALC     | Sometimes | CALC          | Frequently | CALC | Always |
|---------------------|----------|-----------|---------------|------------|------|--------|
| Normal_Weight       | 1.814064 | 3.227314  | 4.195279e-09  |            |      |        |
| Overweight_Level_I  | 2.285957 | 2.959954  | NaN           |            |      |        |
| Overweight_Level_II | 2.473649 | 2.829653  | NaN           |            |      |        |
| Obesity_Type_I      | 2.738216 | 3.347825  | 7.197478e-15  |            |      |        |
| Obesity_Type_II     | 5.876708 | 3.952153  | 5.830840e-27  |            |      |        |
| Obesity_Type_III    | 2.594889 | 6.824172  | 5.863410e-158 |            |      |        |

## Mode of Transportation

|                     | MTRANS       | Bike         | MTRANS   | Motorbike | MTRANS | Public_Transportation | MTRANS | Automobile |
|---------------------|--------------|--------------|----------|-----------|--------|-----------------------|--------|------------|
| Normal_Weight       | 1.729451e+00 | 1.587858e+00 | 3.561331 | 3.803415  |        |                       |        |            |
| Overweight_Level_I  | 1.729451e+00 | 4.537435e+00 | 3.751998 | 4.208075  |        |                       |        |            |
| Overweight_Level_II | NaN          | 3.322124e+00 | 4.640777 | 4.709894  |        |                       |        |            |
| Obesity_Type_I      | 3.194447e-15 | 4.880621e+00 | 5.087815 | 5.970369  |        |                       |        |            |
| Obesity_Type_II     | NaN          | NaN          | 3.379038 | 3.350241  |        |                       |        |            |
| Obesity_Type_III    | 5.743202e-36 | 9.894522e-38 | 6.392352 | 5.781138  |        |                       |        |            |

# K-Nearest Neighbors Analysis



## Key Takeaways:

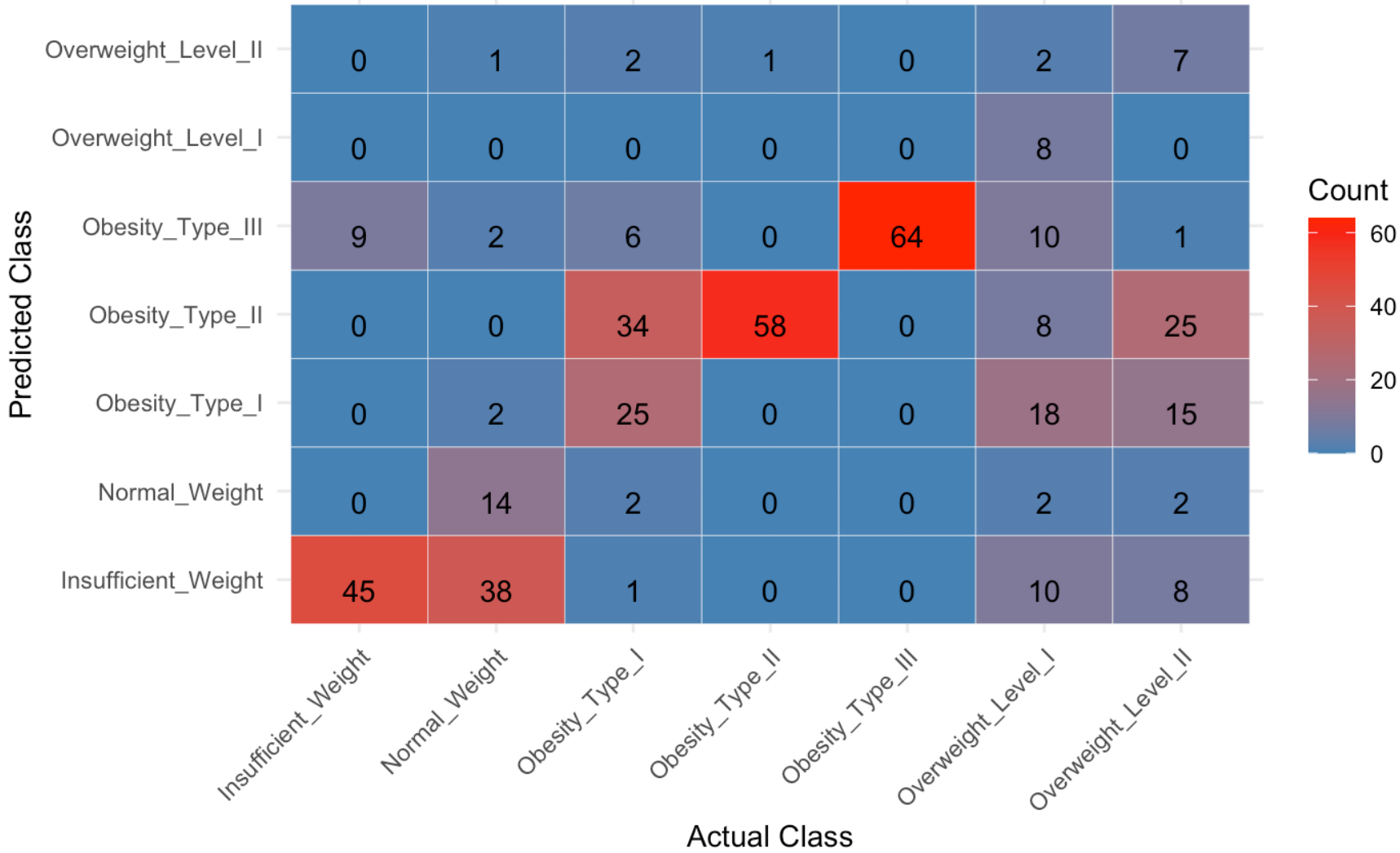
- **Model partition**
  - Train = 80%
  - Test = 20%
- **Similarity-based prediction**
- **Accuracy = 0.8190476**

## Model Performance:

| Scores    | Micro.Average |
|-----------|---------------|
| Precision | 0.962963      |
| Recall    | 0.8387097     |
| F1 Score  | 0.8965517     |

# Naive Bayes Analysis

Confusion Matrix (Naive Bayes)



## Key Takeaways:

- Probabilistic Approach
- Accuracy = 0.5261905

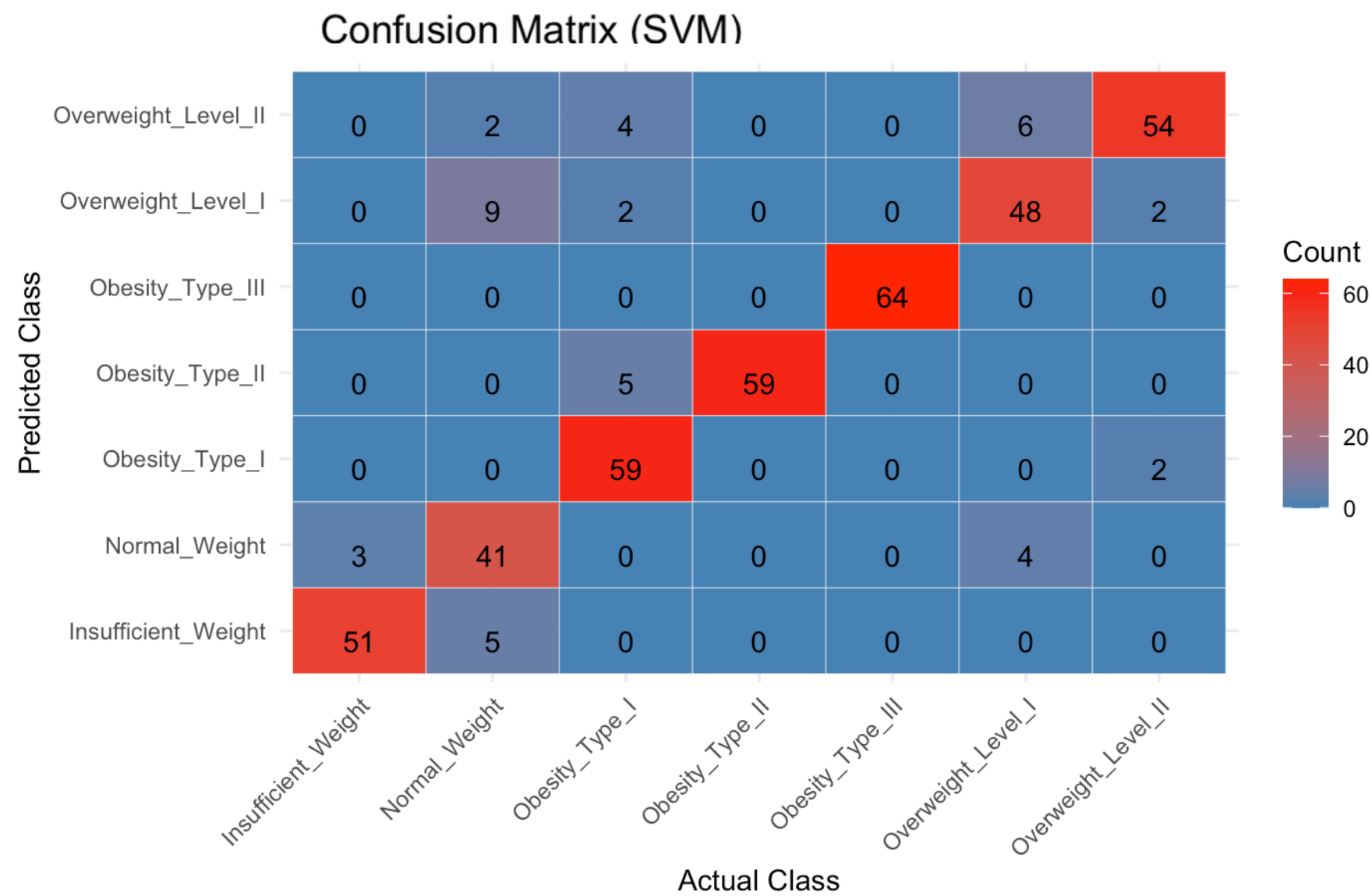
## Model Performance:

| Scores    | Tier 1 | Tier 2 | Tier 3 | Tier 4 | Tier 5 | Tier 6 | Tier 7 |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| Precision | 0.44   | 0.7    | 0.42   | 0.46   | 0.7    | 1      | 0.53   |
| Recall    | 0.83   | 0.25   | 0.35   | 0.98   | 1      | 0.14   | 0.12   |
| F1 Score  | 0.58   | 0.36   | 0.38   | 0.63   | 0.82   | 0.24   | 0.20   |



# SVM Analysis

Group 6



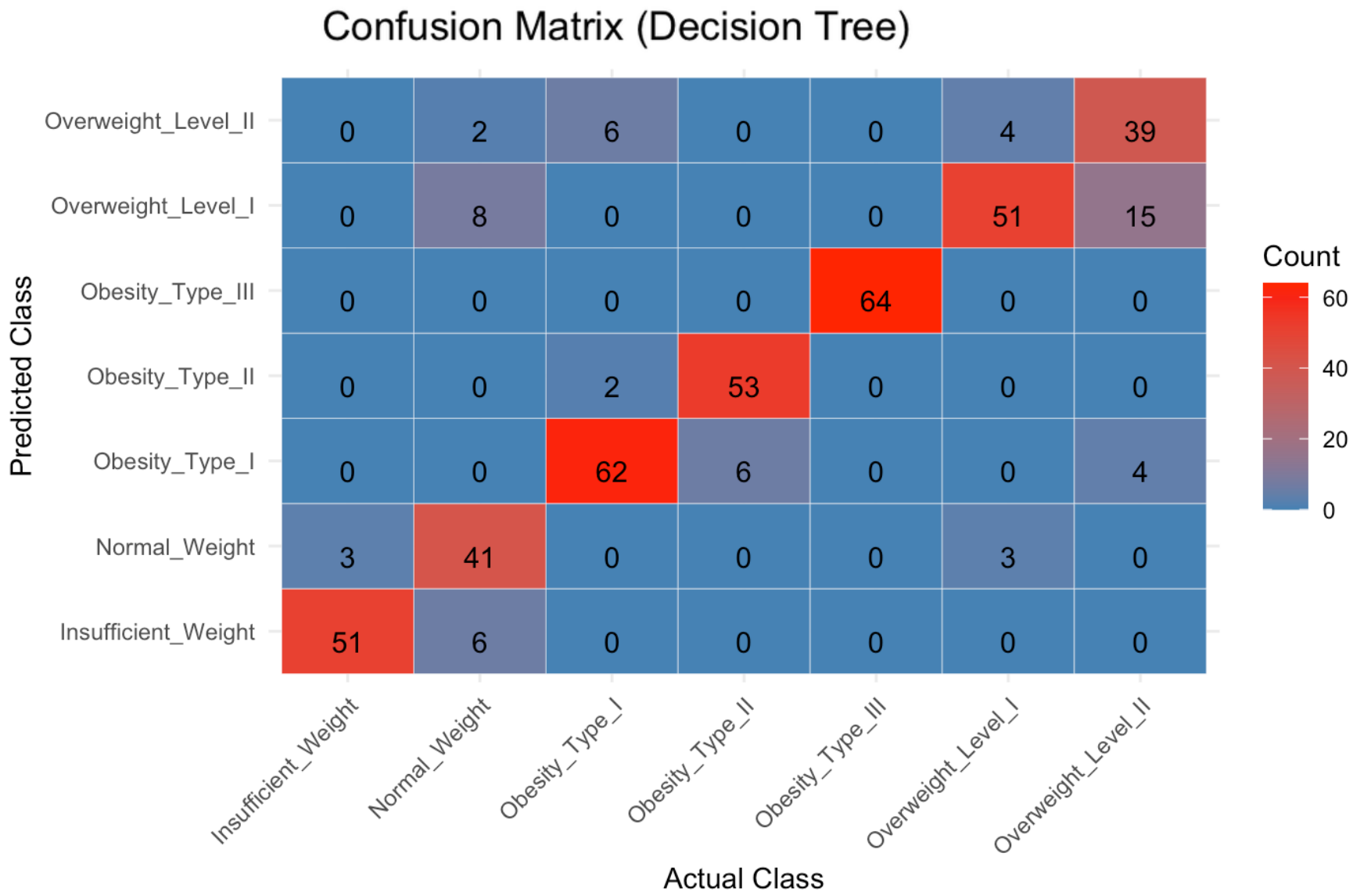
## Key Takeaways:

- **Complex decision boundaries**
- **Accuracy = 0.8952381**

## Model Performance:

| Scores    | Tier 1 | Tier 2 | Tier 3 | Tier 4 | Tier 5 | Tier 6 | Tier 7 |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| Precision | 0.91   | 0.85   | 0.96   | 0.92   | 1      | 0.79   | 0.81   |
| Recall    | 0.94   | 0.72   | 0.84   | 1      | 1      | 0.83   | 0.93   |
| F1 Score  | 0.93   | 0.78   | 0.90   | 0.96   | 1      | 0.81   | 0.87   |

# Decision Tree Analysis



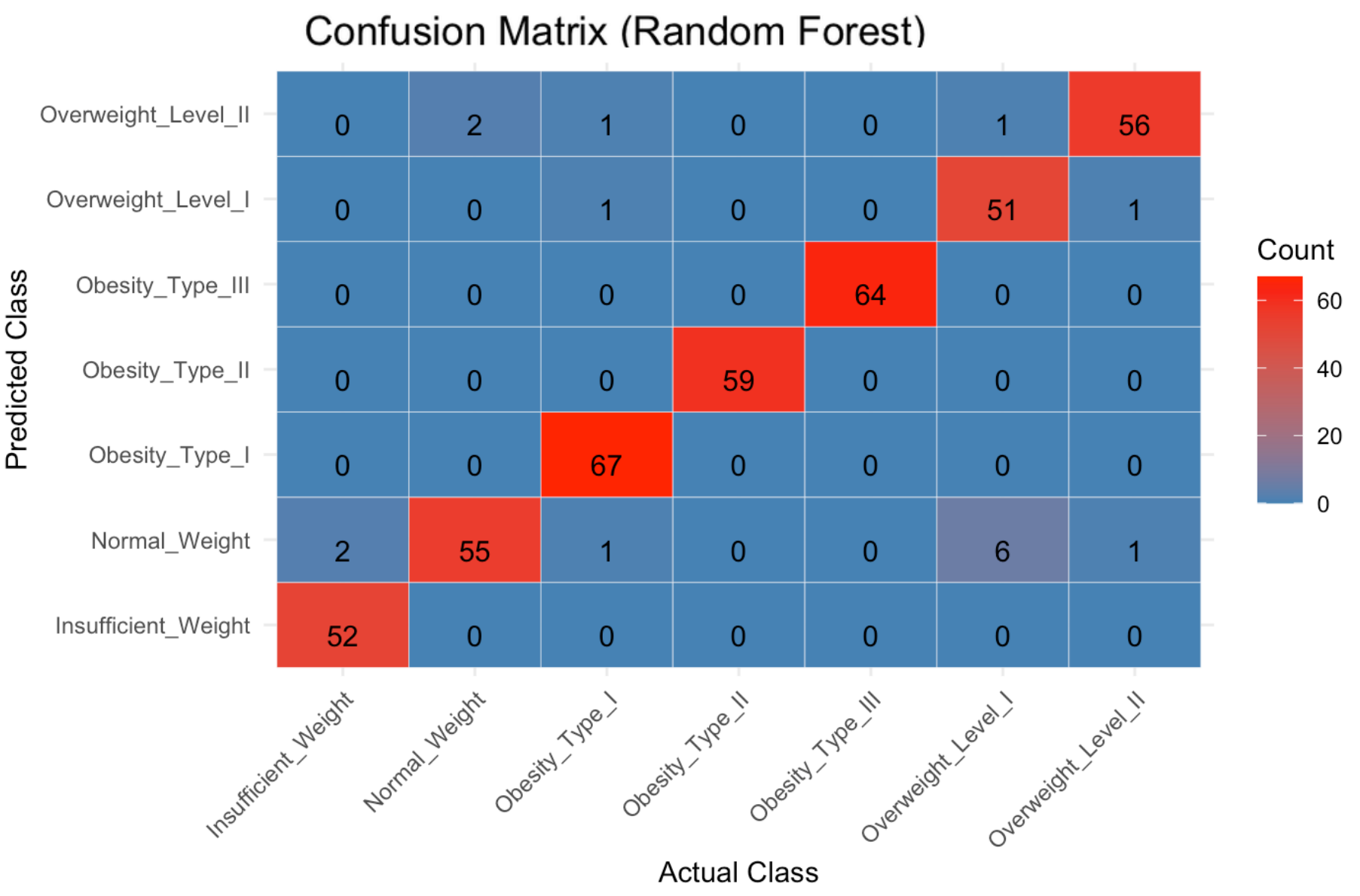
## Key Takeaways:

- Easy Interpretation
- **Accuracy** = 0.8595238
- Possibility of **Overfitting**

## Model Performance:

| Scores    | Tier 1 | Tier 2 | Tier 3 | Tier 4 | Tier 5 | Tier 6 | Tier 7 |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| Precision | 0.89   | 0.87   | 0.86   | 0.96   | 1      | 0.69   | 0.76   |
| Recall    | 0.94   | 0.72   | 0.89   | 0.90   | 1      | 0.88   | 0.67   |
| F1 Score  | 0.92   | 0.79   | 0.87   | 0.93   | 1      | 0.77   | 0.72   |

# Random Forest



## Key Takeaways:

- Handling **complexity** and **overfitting**
- **Accuracy** = 0.9619048

## Model Performance:

| Scores    | Tier 1 | Tier 2 | Tier 3 | Tier 4 | Tier 5 | Tier 6 | Tier 7 |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| Precision | 1      | 0.85   | 1      | 1      | 1      | 0.96   | 0.93   |
| Recall    | 0.96   | 0.96   | 0.95   | 1      | 1      | 0.88   | 0.97   |
| F1 Score  | 0.98   | 0.90   | 0.98   | 1      | 1      | 0.92   | 0.95   |



# Variable Importance Table

|                                    | Insufficient_Weight | Normal_Weight | Obesity_Type_I | Obesity_Type_II | Obesity_Type_III | Overweight_Level_I | Overweight_Level_II | MeanDecreaseAccuracy | MeanDecreaseGini |
|------------------------------------|---------------------|---------------|----------------|-----------------|------------------|--------------------|---------------------|----------------------|------------------|
| Gender.Female                      | 18.003261           | 7.60622374    | 18.641438      | 17.9090194      | 19.744122        | 17.354181          | 17.4427610          | 22.166797            | 50.3666448       |
| Gender.Male                        | 16.993355           | 9.01709509    | 19.759741      | 18.4999213      | 19.556559        | 16.142623          | 16.9795389          | 22.439608            | 52.4813644       |
| Age                                | 34.880107           | 9.25892948    | 36.174965      | 31.6586884      | 13.375984        | 34.783943          | 33.8920276          | 43.804390            | 124.9994899      |
| Height                             | 25.568005           | 18.84666642   | 36.624444      | 17.4955857      | 9.923309         | 34.827523          | 34.3439269          | 43.035569            | 109.6431761      |
| Weight                             | 67.934212           | 51.94611205   | 60.610805      | 71.2176996      | 44.615821        | 48.742049          | 50.2216814          | 80.381385            | 390.6006529      |
| family_history_with_overweight.no  | 20.004804           | 9.26666969    | 18.506932      | 13.3239748      | 12.657590        | 17.414121          | 17.1825691          | 21.642764            | 30.7848658       |
| family_history_with_overweight.yes | 18.274134           | 8.32281320    | 16.462077      | 12.8426960      | 12.292250        | 15.724462          | 14.8157279          | 19.248172            | 28.0431643       |
| FAVC.no                            | 10.756355           | 1.07152922    | 15.252472      | 7.4977988       | 7.517012         | 11.900162          | 19.0192120          | 20.065393            | 16.4860749       |
| FAVC.yes                           | 10.790386           | 2.40849641    | 14.454256      | 7.0847803       | 8.079190         | 13.761651          | 20.3614980          | 20.363651            | 17.1664249       |
| FCVC                               | 23.289766           | 11.69619874   | 32.825840      | 24.3615909      | 24.973007        | 26.464250          | 26.0316299          | 33.841351            | 113.0823447      |
| NCP                                | 25.353015           | 7.31196662    | 25.599636      | 21.6104901      | 14.299904        | 27.108660          | 26.2953411          | 34.054383            | 69.0150119       |
| CAEC.Always                        | 3.971884            | 5.25458986    | 2.574971       | 2.5410651       | 2.147092         | 1.969569           | 1.3632699           | 7.457542             | 4.1732922        |
| CAEC.Frequently                    | 24.145026           | 2.84096576    | 14.656448      | 9.9364405       | 8.615937         | 14.071507          | 10.4344688          | 22.949735            | 22.2544268       |
| CAEC.no                            | 5.110066            | -3.83529051   | 4.404233       | 3.4414995       | 1.583788         | 14.933967          | 2.6484284           | 14.223325            | 6.1688721        |
| CAEC.Sometimes                     | 19.541529           | 9.98253611    | 16.569783      | 11.7482988      | 11.665911        | 15.462014          | 14.0002601          | 23.029888            | 27.2196416       |
| SMOKE.no                           | 3.192800            | -0.67632992   | 2.385053       | 1.0814660       | 3.009892         | 3.110074           | 2.3784169           | 4.427804             | 2.8729939        |
| SMOKE.yes                          | 3.272485            | 0.57634542    | 2.117369       | 0.7863754       | 3.491194         | 3.382698           | 2.0381386           | 5.642659             | 2.7491837        |
| CH2O                               | 17.278048           | 14.84323096   | 25.396029      | 23.6293311      | 7.274692         | 21.841430          | 18.5584172          | 37.162350            | 58.2361369       |
| SCC.no                             | 4.068299            | -2.30664424   | 5.397666       | 3.6014654       | 3.716449         | 13.825764          | 6.1504452           | 13.720160            | 6.9864651        |
| SCC.yes                            | 4.560923            | -3.15721654   | 5.300215       | 3.1122291       | 4.473467         | 13.836232          | 4.8446225           | 12.966461            | 7.2189122        |
| FAF                                | 19.402038           | 9.09944674    | 22.999842      | 20.3604531      | 9.954861         | 22.076762          | 16.7767060          | 32.465681            | 57.3758010       |
| TUE                                | 26.135949           | 11.42681814   | 23.873489      | 17.2863824      | 14.124842        | 18.667840          | 23.6594425          | 29.818390            | 62.5779492       |
| CALC.Always                        | 0.000000            | 0.00000000    | 0.000000       | 0.0000000       | 0.000000         | 0.000000           | 0.0000000           | 0.000000             | 0.0000000        |
| CALC.Frequently                    | 4.813216            | -2.36075912   | 8.703720       | 4.8107664       | 2.651261         | 7.151018           | 5.2809901           | 10.956701            | 5.8944940        |
| CALC.no                            | 17.590562           | -1.57720526   | 17.720477      | 13.3127643      | 11.540718        | 18.211232          | 18.9608466          | 22.267368            | 27.5029376       |
| CALC.Sometimes                     | 17.481260           | 1.21304138    | 19.472071      | 14.5966402      | 13.418356        | 22.308033          | 20.0893505          | 23.567616            | 29.8163886       |
| MTRANS.Automobile                  | 17.295471           | -0.72615946   | 15.551812      | 8.4450407       | 7.882614         | 16.885564          | 16.4449524          | 22.804124            | 19.4189724       |
| MTRANS.Bike                        | 1.735127            | 1.99548286    | 1.735237       | 1.0010015       | 0.000000         | -1.001002          | 0.0000000           | 2.872482             | 0.9112008        |
| MTRANS.Motorbike                   | 1.416457            | -0.06375631   | 1.001002       | 1.4170293       | 1.416502         | 0.000000           | -0.1042432          | 1.423692             | 0.8914170        |
| MTRANS.Public_Transportation       | 17.786655           | 4.34439932    | 15.971001      | 10.1823972      | 9.897790         | 17.825494          | 16.7224260          | 23.050893            | 22.1465193       |
| MTRANS.Walking                     | 3.167854            | 0.52029928    | 3.022157       | 4.4141583       | 2.685434         | 1.770577           | 4.3973267           | 6.243278             | 4.5638339        |



|                                    | MeanDecreaseAccuracy | MeanDecreaseGini |
|------------------------------------|----------------------|------------------|
| Gender.Female                      | 22.166797            | 50.3666448       |
| Gender.Male                        | 22.439608            | 52.4813644       |
| Age                                | 43.804390            | 124.9994899      |
| Height                             | 43.035569            | 109.6431761      |
| Weight                             | 80.381385            | 390.6006529      |
| family_history_with_overweight.no  | 21.642764            | 30.7848658       |
| family_history_with_overweight.yes | 19.248172            | 28.0431643       |
| FAVC.no                            | 20.065393            | 16.4860749       |
| FAVC.yes                           | 20.363651            | 17.1664249       |
| FCVC                               | 33.841351            | 113.0823447      |
| NCP                                | 34.054383            | 69.0150119       |
| CAEC.Always                        | 7.457542             | 4.1732922        |
| CAEC.Frequently                    | 22.949735            | 22.2544268       |
| CAEC.no                            | 14.223325            | 6.1688721        |
| CAEC.Sometimes                     | 23.029888            | 27.2196416       |
| SMOKE.no                           | 4.427804             | 2.8729939        |
| SMOKE.yes                          | 5.642659             | 2.7491837        |
| CH2O                               | 37.162350            | 58.2361369       |
| SCC.no                             | 13.720160            | 6.9864651        |
| SCC.yes                            | 12.966461            | 7.2189122        |
| FAF                                | 32.465681            | 57.3758010       |
| TUE                                | 29.818390            | 62.5779492       |
| CALC.Always                        | 0.000000             | 0.0000000        |
| CALC.Frequently                    | 10.956701            | 5.8944940        |
| CALC.no                            | 22.267368            | 27.5029376       |
| CALC.Sometimes                     | 23.567616            | 29.8163886       |
| MTRANS.Automobile                  | 22.804124            | 19.4189724       |
| MTRANS.Bike                        | 2.872482             | 0.9112008        |
| MTRANS.Motorbike                   | 1.423692             | 0.8914170        |
| MTRANS.Public_Transportation       | 23.050893            | 22.1465193       |
| MTRANS.Walking                     | 6.243278             | 4.5638339        |

# Thank you

- **STAT3622 Final Project Presentation by Group 6**  
Soo-ah Kim (3035661061) & Dongjun Yeom (3035666463)