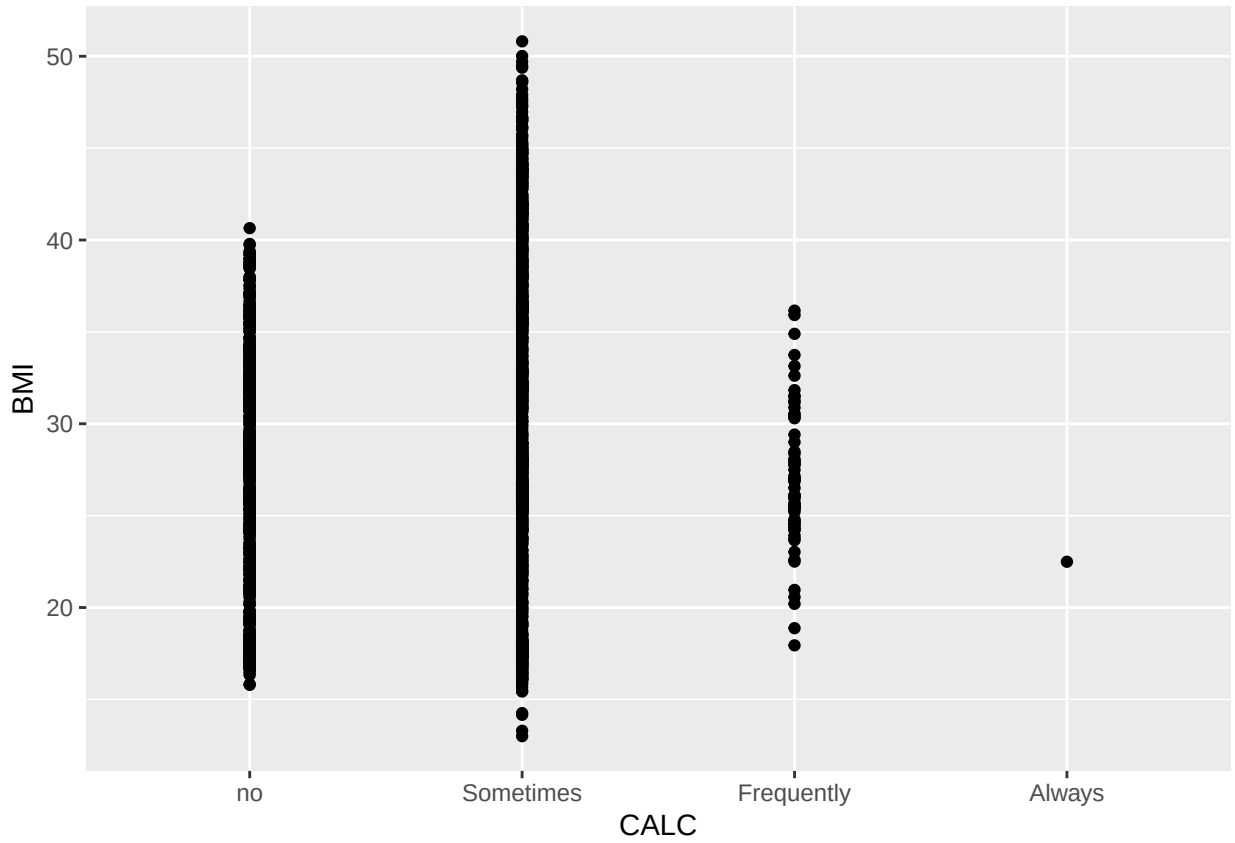
DATA 2010 Project

Ethan Robson, 07891904, A02

2024-03-08

```
mapping2 = function(x){
  new_vars = c()

  for(i in 1:length(x)){


    if(x[i] == "no"){
      new_vars = c(new_vars, 1)
    } else if(x[i] == "Sometimes"){
      new_vars = c(new_vars, 2)
    } else if(x[i] == "Frequently"){
      new_vars = c(new_vars, 3)
    } else {
      new_vars = c(new_vars, 4)
    }
  }
  return(new_vars)

}
factored = data %>% mutate(BMI = Weight / (Height^2),
                           CALC = factor(CALC, levels = c("no", "Sometimes",
                                                          "Frequently", "Always")),
                           CALCN = mapping2(CALC))

factored %>% ggplot(aes(x = CALC, y = BMI)) +
  geom_point()
```




```
cor(factored$CALCN, factored$BMI, method = "spearman")
```

```
## [1] 0.1610393
```

```
unique(factored$CALC)
```

```
## [1] no         Sometimes  Frequently Always
## Levels: no Sometimes Frequently Always
```

```
str(factored)
```

```
## 'data.frame':    2111 obs. of  19 variables:
##  $ Gender                         : chr  "Female" "Female" "Male" "Male" ...
##  $ Age                            : num  21 21 23 27 22 29 23 22 24 22 ...
##  $ Height                         : num  1.62 1.52 1.8 1.8 1.78 1.62 1.5 1.64 1.78 1.72 ...
##  $ Weight                         : num  64 56 77 87 89.8 53 55 53 64 68 ...
##  $ family_history_with_overweight: chr  "yes" "yes" "yes" "no" ...
##  $ FAVC                           : chr  "no" "no" "no" "no" ...
##  $ FCVC                           : num  2 3 2 3 2 2 3 2 3 2 ...
##  $ NCP                            : num  3 3 3 3 1 3 3 3 3 3 ...
##  $ CAEC                           : chr  "Sometimes" "Sometimes" "Sometimes" "Sometimes" ...
##  $ SMOKE                          : chr  "no" "yes" "no" "no" ...
##  $ CH2O                           : num  2 3 2 2 2 2 2 2 2 2 ...
##  $ SCC                            : chr  "no" "yes" "no" "no" ...
```

```
##  $ FAF                           : num  0 3 2 2 0 0 1 3 1 1 ...
##  $ TUE                           : num  1 0 1 0 0 0 0 0 1 1 ...
##  $ CALC                          : Factor w/ 4 levels "no","Sometimes",..: 1 2 3 3 2 2 2 2 3 1 ...
##  $ MTRANS                        : chr  "Public_Transportation" "Public_Transportation" "Public_Trans
##  $ NObeyesdad                    : chr  "Normal_Weight" "Normal_Weight" "Normal_Weight" "Overweight_
##  $ BMI                           : num  24.4 24.2 23.8 26.9 28.3 ...
##  $ CALCN                         : num  1 2 3 3 2 2 2 2 3 1 ...
```

```
str(data)
```

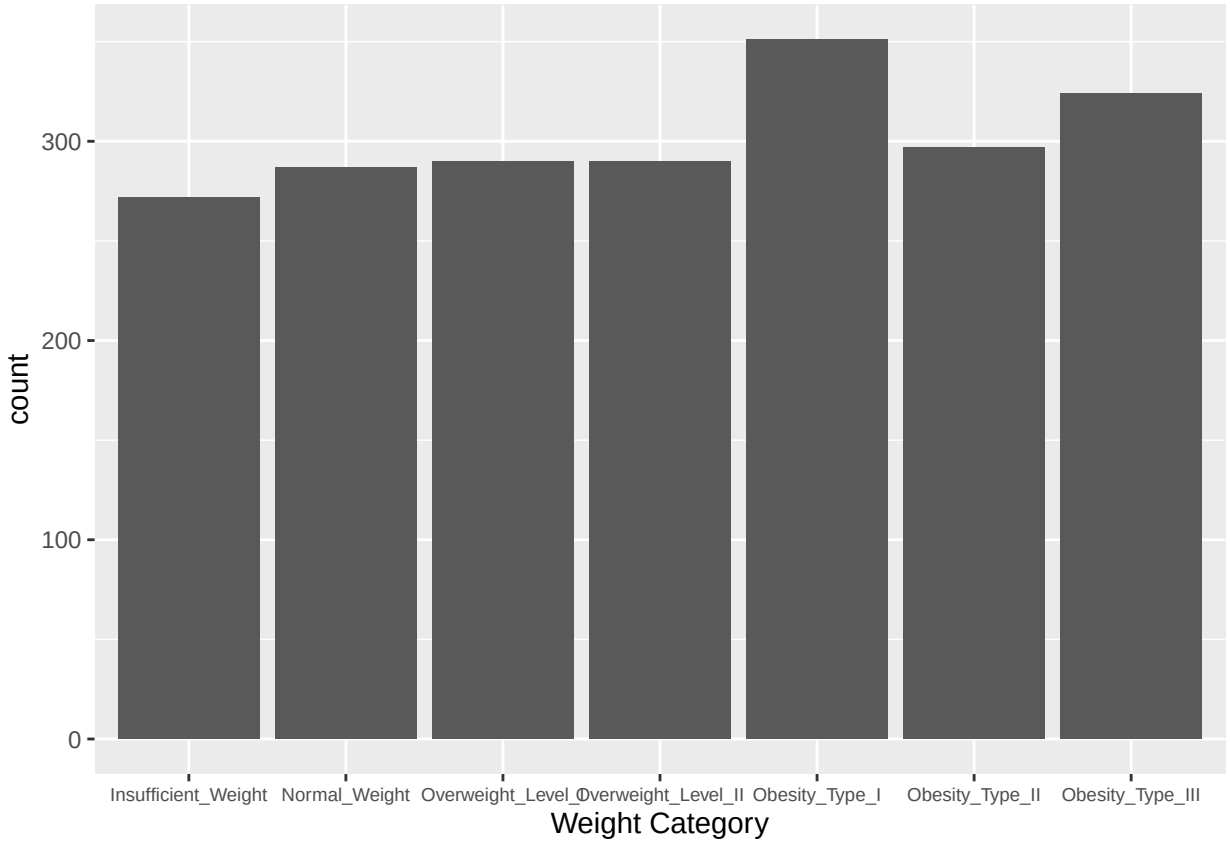
```
## 'data.frame':    2111 obs. of  17 variables:
##  $ Gender                        : chr  "Female" "Female" "Male" "Male" ...
##  $ Age                           : num  21 21 23 27 22 29 23 22 24 22 ...
##  $ Height                        : num  1.62 1.52 1.8 1.8 1.78 1.62 1.5 1.64 1.78 1.72 ...
##  $ Weight                        : num  64 56 77 87 89.8 53 55 53 64 68 ...
##  $ family_history_with_overweight: chr  "yes" "yes" "yes" "no" ...
##  $ FAVC                          : chr  "no" "no" "no" "no" ...
##  $ FCVC                          : num  2 3 2 3 2 2 3 2 3 2 ...
##  $ NCP                           : num  3 3 3 3 1 3 3 3 3 3 ...
##  $ CAEC                          : chr  "Sometimes" "Sometimes" "Sometimes" "Sometimes" ...
##  $ SMOKE                         : chr  "no" "yes" "no" "no" ...
##  $ CH2O                          : num  2 3 2 2 2 2 2 2 2 2 ...
##  $ SCC                           : chr  "no" "yes" "no" "no" ...
##  $ FAF                           : num  0 3 2 2 0 0 1 3 1 1 ...
##  $ TUE                           : num  1 0 1 0 0 0 0 0 1 1 ...
##  $ CALC                          : chr  "no" "Sometimes" "Frequently" "Frequently" ...
##  $ MTRANS                        : chr  "Public_Transportation" "Public_Transportation" "Public_Trans
##  $ NObeyesdad                    : chr  "Normal_Weight" "Normal_Weight" "Normal_Weight" "Overweight_
```

```
unique(data$NObeyesdad)
```

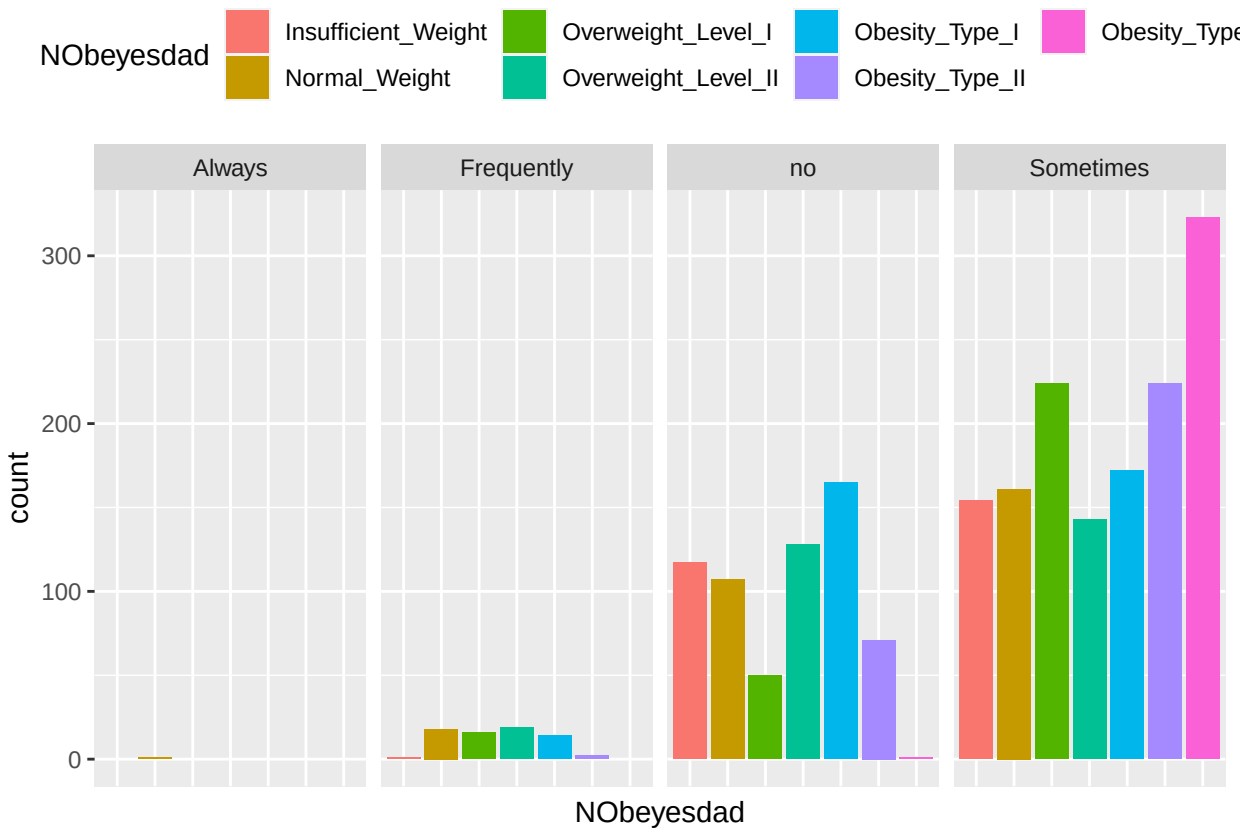
```
## [1] "Normal_Weight"       "Overweight_Level_I"  "Overweight_Level_II"
## [4] "Obesity_Type_I"      "Insufficient_Weight" "Obesity_Type_II"
## [7] "Obesity_Type_III"
```

```
mapping = function(x){
  new_vars = c()

  for(i in 1:length(x)){


    if(x[i] == "Insufficient_Weight" || x[i] == "Normal_Weight"){
      new_vars = c(new_vars, "NON")
    } else if(x[i] == "Overweight_Level_I" || x[i] == "Overweight_Level_II"){
      new_vars = c(new_vars, "OVR")
    } else {
      new_vars = c(new_vars, "OBE")
    }
  }
  return(new_vars)
}
```

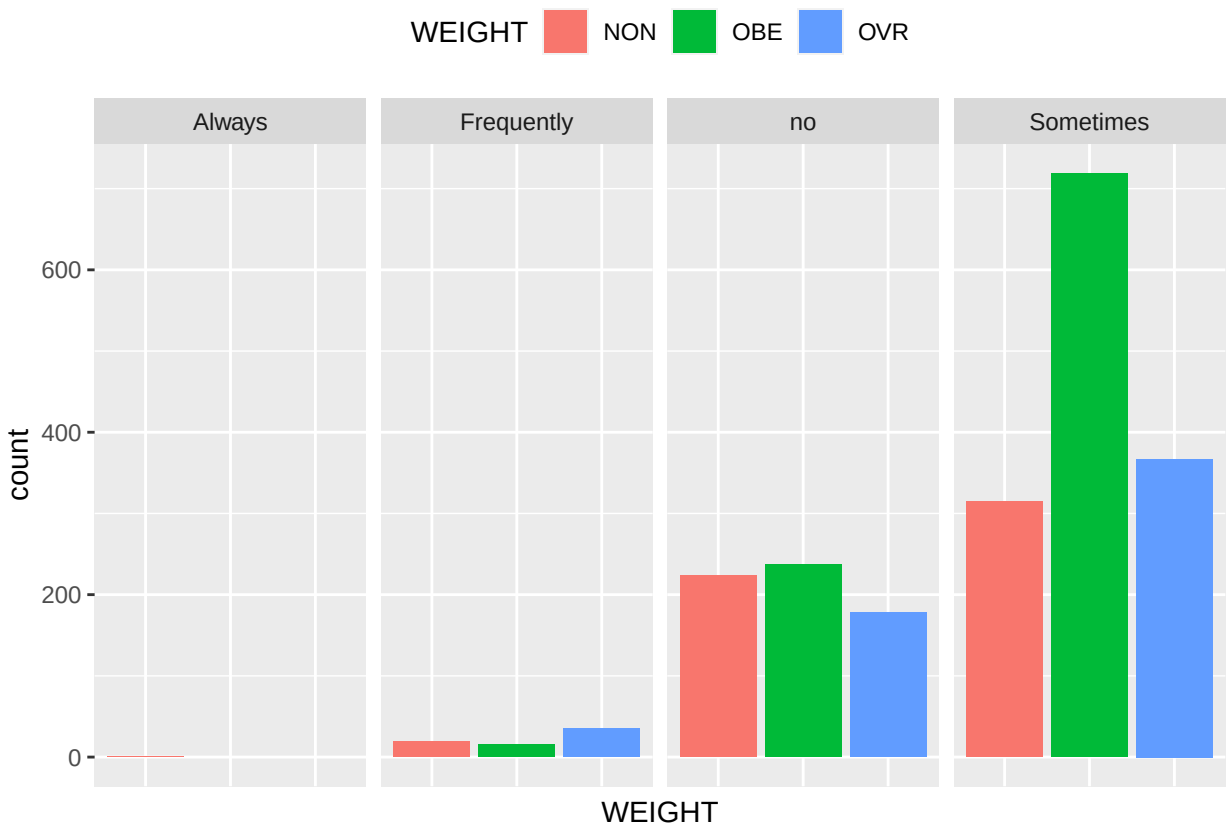
```r
data = data %>%
  mutate(NObeyesdad = factor(NObeyesdad, levels = c("Insufficient_Weight", "Normal_Weight",
                                                    "Overweight_Level_I", "Overweight_Level_II",
                                                    "Obesity_Type_I", "Obesity_Type_II",
                                                    "Obesity_Type_III")),
         CALC = factor(CALC),
         WEIGHT = mapping(NObeyesdad))

data %>%
  ggplot(aes(x = NObeyesdad)) +
  geom_bar() +
  theme(axis.text.x=element_text(size=7)) +
  xlab("Weight Category")
```



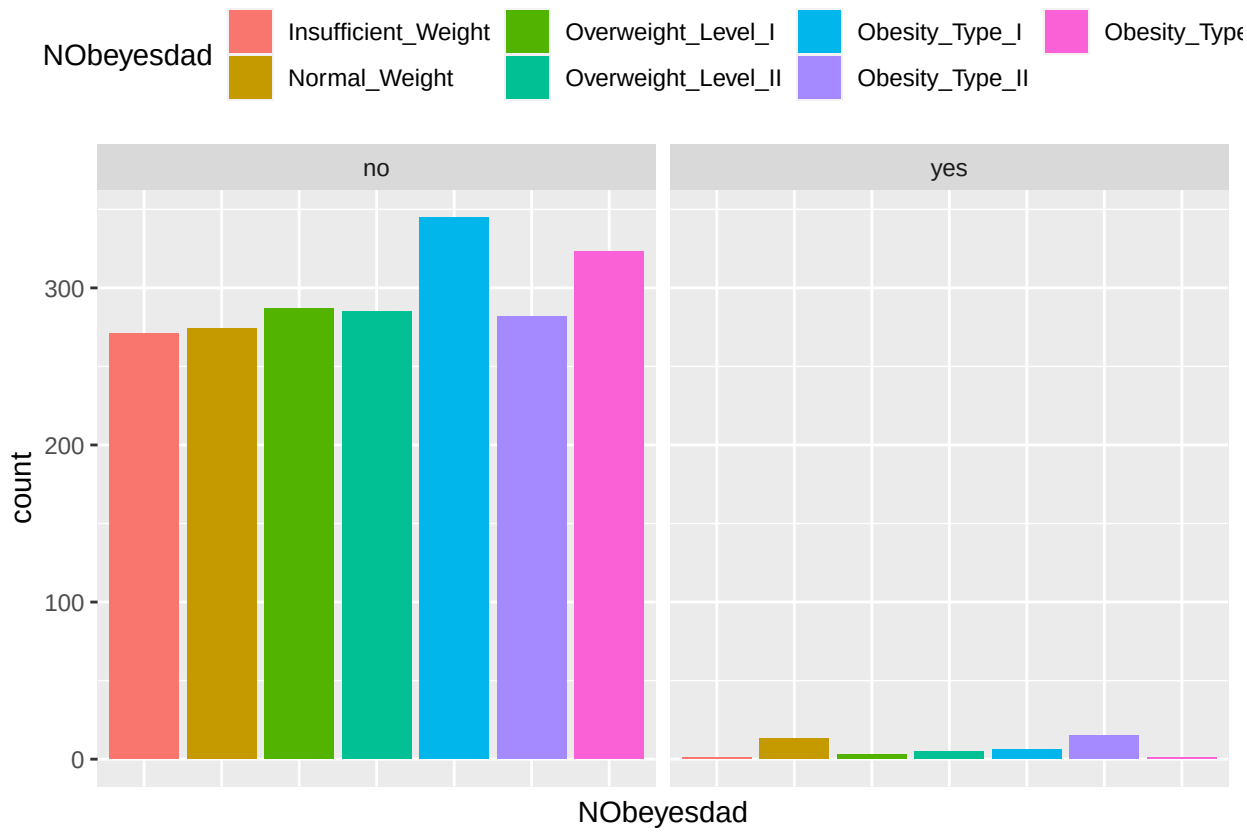

```r
data %>%
  ggplot(aes(x = NObeyesdad, fill = NObeyesdad)) +
  geom_bar() +
  theme(legend.position = "top",
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
  facet_grid(~ CALC)
```



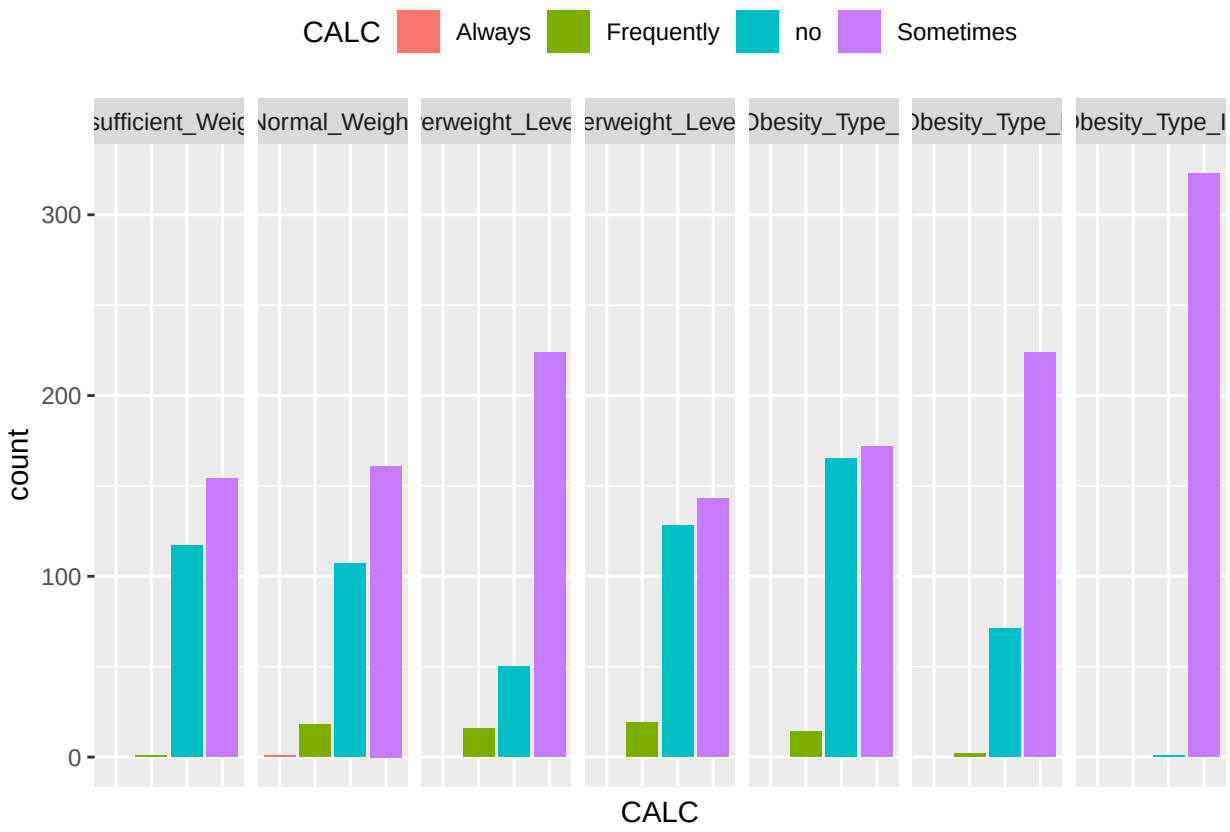

```
data %>%
  ggplot(aes(x = WEIGHT, fill = WEIGHT)) +
  geom_bar() +
  theme(legend.position = "top",
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
  facet_grid(~ CALC)
```




```
data %>%
  ggplot(aes(x = NObeyesdad, fill = NObeyesdad)) +
  geom_bar() +
  theme(legend.position = "top",
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
  facet_grid(~ SMOKE)
```




```
data %>%
  ggplot(aes(x = CALC, fill = CALC)) +
  geom_bar() +
  theme(legend.position = "top",
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
  facet_grid(~ NObeyesdad)
```


CALC
Always
Frequently
no
Sometimes
sufficient_Weig
Normal_Weigh
erweight_Leve
erweight_Leve
Obesity_Type_
Obesity_Type_
besity_Type_I
300
200
100
0
count
CALC