

# kruskal-wallis

```
data <- diabeted_multiclass %>%  
  select(Diabetes_012,  
         BMI)  
  
data$Diabetes_012 <- factor(data$Diabetes_012,  
                           levels = c(0, 1, 2),  
                           labels = c("No Diabetes", "Prediabetes", "Diabetes"))  
  
glimpse(data)
```

Rows: 253,680

Columns: 2

\$ Diabetes\_012 <fct> No Diabetes, No Diabetes, No Diabetes, No Diabetes, No Di~

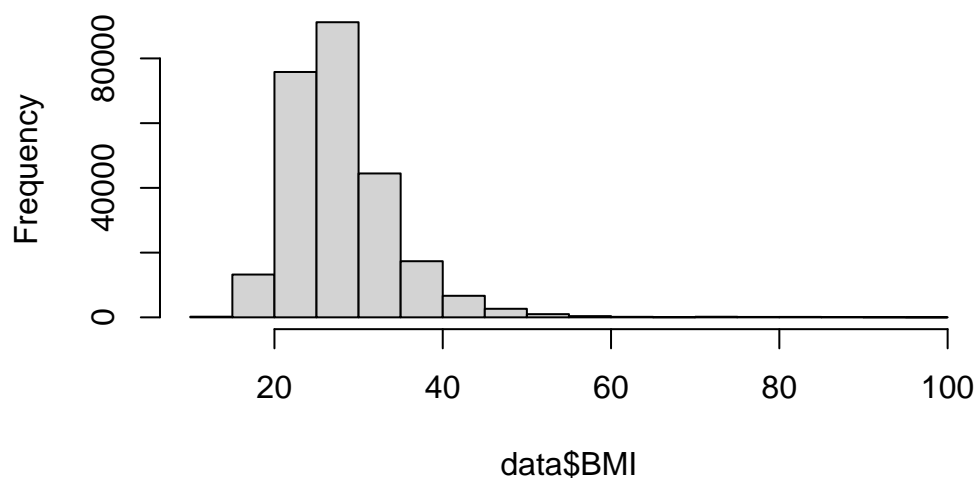
\$ BMI <dbl> 40, 25, 28, 27, 24, 25, 30, 25, 30, 24, 25, 34, 26, 28, 3~

```
summary(data$BMI)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12.00	24.00	27.00	28.38	31.00	98.00

```
hist(data$BMI)
```

## Histogram of data\$BMI

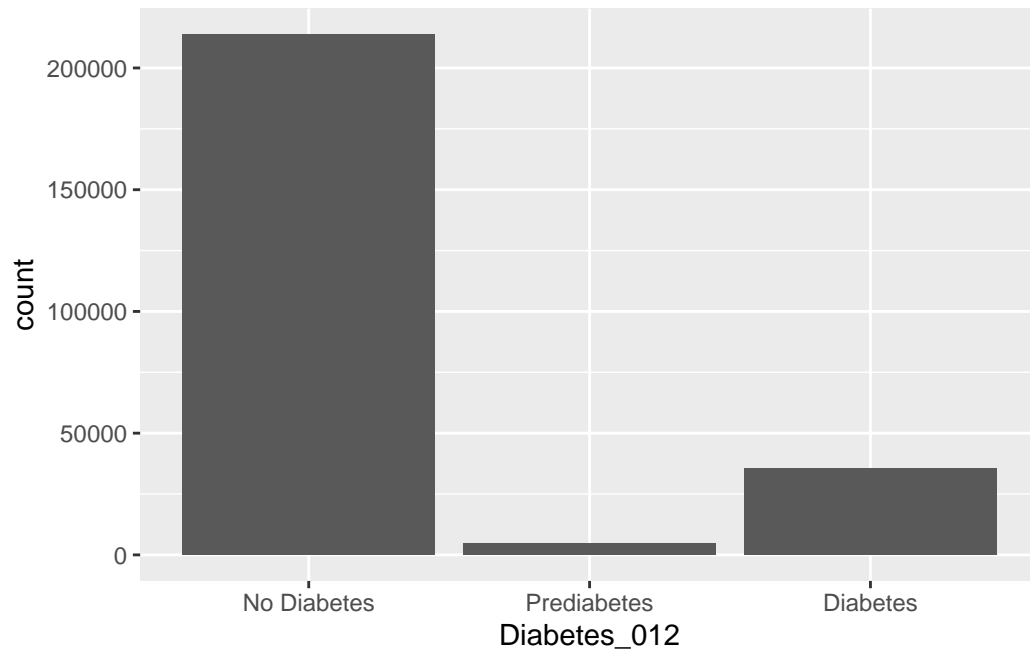


```
p_0 <- data %>%  
  filter(Diabetes_012 == "No Diabetes") %>%  
  ggplot()+  
  geom_histogram(aes(x = BMI))
```

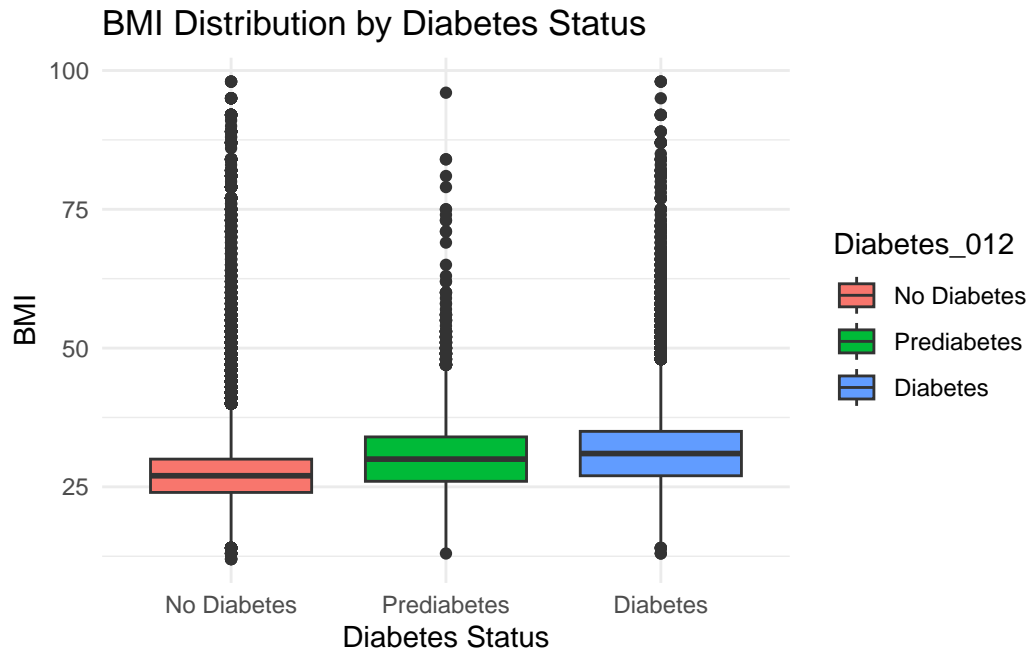
```
p_1 <- data %>%  
  filter(Diabetes_012 == "Prediabetes") %>%  
  ggplot()+  
  geom_histogram(aes(x = BMI))
```

```
p_2 <- data %>%  
  filter(Diabetes_012 == "Diabetes") %>%  
  ggplot()+  
  geom_histogram(aes(x = BMI))
```

```
data %>%  
  ggplot()+  
  geom_bar(aes(x = Diabetes_012))
```



```
ggplot(data, aes(x = Diabetes_012, y = BMI, fill = Diabetes_012)) +  
  geom_boxplot() +  
  labs(title = "BMI Distribution by Diabetes Status",  
        x = "Diabetes Status",  
        y = "BMI") +  
  theme_minimal()
```



```
kruskal.test(BMI ~ Diabetes_012, data = data)
```

Kruskal-Wallis rank sum test

data: BMI by Diabetes\_012

Kruskal-Wallis chi-squared = 14131, df = 2, p-value < 2.2e-16

```
pairwise.wilcox.test(data$BMI, data$Diabetes_012, p.adjust.method = "bonferroni")
```

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: data\$BMI and data\$Diabetes\_012

	No Diabetes	Prediabetes
Prediabetes	<2e-16	-
Diabetes	<2e-16	<2e-16

P value adjustment method: bonferroni