Final Project- Marijuana consumption effect on Physical and Mental health

Murtaza Khasamwala

2024-12-13

Introduction:

The consumption of marijuana has become a topic of increasing interest due to its growing use for both recreational and medicinal purposes. However, its effects on various aspects of health remain a subject of debate. Understanding the impact of marijuana consumption on health parameters, particularly physical and mental health, is critical for developing public health guidelines and addressing unmet health needs. To base the policies on practical needs, such analysis aims to find the potentials benefits or hazards on overall health caused by the consumption of Marijuana.

Method:

Data Source Analysis was conducted using data from the National Health and Nutrition Examination Survey (NHANES).

Study Parameters The study focuses on people of age greater than 18, including all ethnicities and gender. The study observes the mean BMI and mental health condition of people categorized as Yes and No in Marijuana consumption column: categorising in two major categories: Smoker and Non-Smoker.

Statistical Analysis Descriptive statistics like mean, median, standard deviation is calculated for study parameter BMI (body – mass index) of smokers and non-smokers, and proportion of prevalence of depression and anxiety between smoker and non-smoker groups.

Hypothesis It is hypothesized that the mean BMI of smoker and non-smokers is not different and there is less prevalence of depression in smokers than the non-smokers

Physical Health

Null Hypothesis (Ho): Marijuana consumption has an effect on BMI

Alternate Hypthesis (Ha): Marijuana consumption has not effect on BMI

Mental Health

Null Hypothesis (Ho): Marijuana consumption has no signficant impact on the mental health of an individual

Alternate Hypthesis (Ha): Marijuana consumption has an effect on the mental health of an individual

Data Processing

BMI statistics Table:

```
bmi_stats <- data %>%
  group_by(RegularMarij, Gender, Race1, AgeDecade) %>%
  summarise(mean_BMI = mean(BMI, na.rm = TRUE),
            count = n())
```

```
## `summarise()` has grouped output by 'RegularMarij', 'Gender', 'Race1'. You can
## override using the `.groups` argument.
```

```
print(bmi_stats)
```

```
## # A tibble: 89 × 6
## # Groups:   RegularMarij, Gender, Race1 [20]
##    RegularMarij Gender Race1    AgeDecade mean_BMI count
##    <fct>        <fct>  <chr>    <chr>        <dbl> <int>
##  1 No           female Black    10-19         29.8     1
##  2 No           female Black    20-29         32.5    21
##  3 No           female Black    30-39         36.4    11
##  4 No           female Black    40-49         37.4     9
##  5 No           female Black    50-59         32.6    11
##  6 No           female Hispanic 20-29         26.9     7
##  7 No           female Hispanic 30-39         28.6     8
##  8 No           female Hispanic 40-49         28.4     4
##  9 No           female Mexican  10-19         28.5     1
## 10 No           female Mexican  20-29         29.3    18
## # ℹ 79 more rows
```

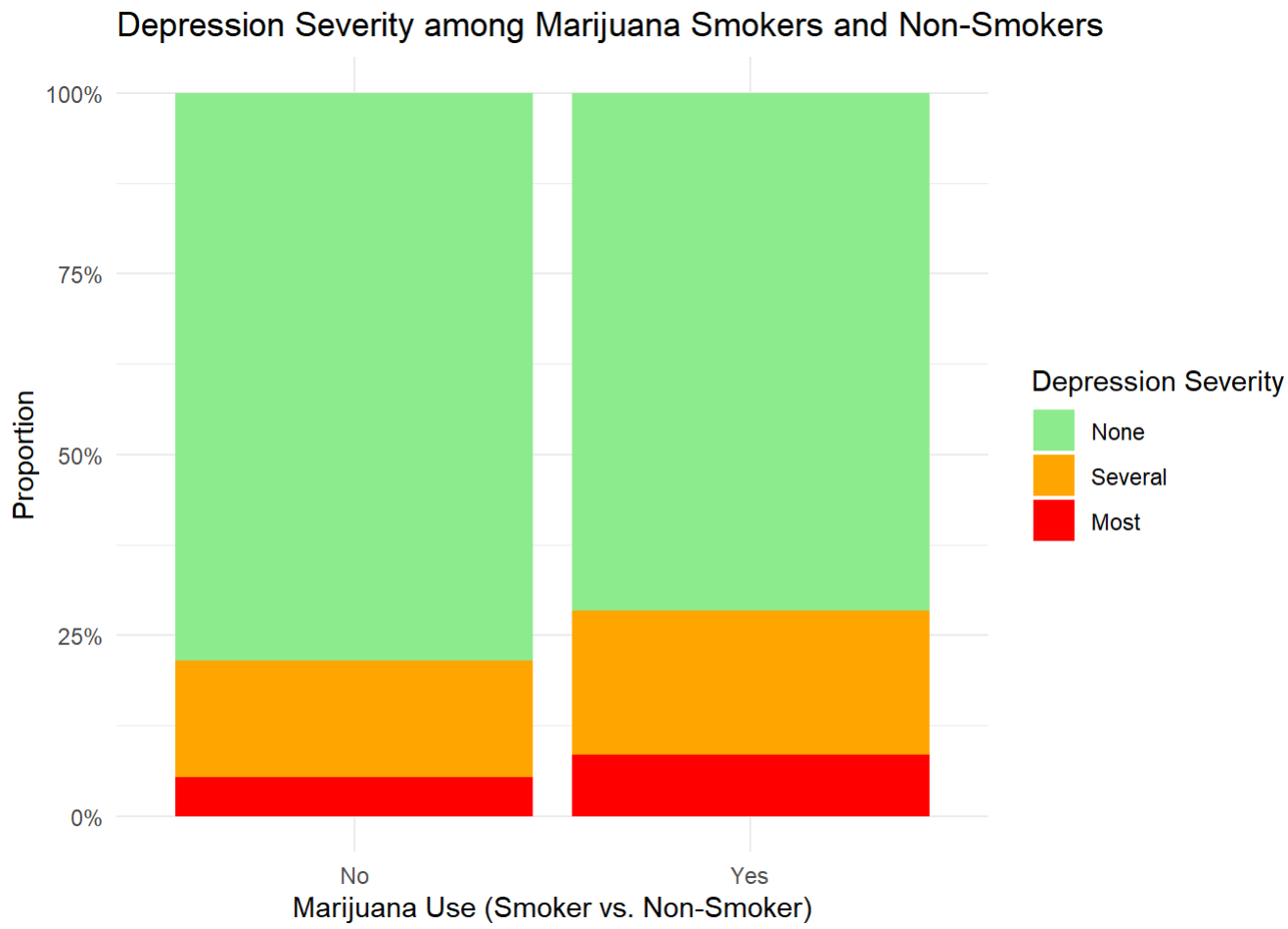
```
vars <- c('BMI', 'Gender', 'Race1', 'AgeDecade')
# Define factor variables
factorVars <- c('Gender', 'Race1', 'AgeDecade', 'RegularMarij')
# Create the summary table (stratified by RegularMarij, i.e., marijuana smokers and non-smoke
rs)
table1 <- CreateTableOne(vars = vars, strata = "RegularMarij", data = data, factorVars = fact
orVars)
print(table1)
```

```
##                    Stratified by RegularMarij
##                     No            Yes           p      test
##   n                  1397          1231
##   BMI (mean (SD))   28.50 (6.37)  28.45 (6.62)   0.833
##   Gender = male (%)   700 (50.1)    767 (62.3)  <0.001
##   Race1 (%)                                     <0.001
##      Black            111 ( 7.9)    158 (12.8)
##      Hispanic          44 ( 3.1)     50 ( 4.1)
##      Mexican           93 ( 6.7)     62 ( 5.0)
##      Other             65 ( 4.7)     70 ( 5.7)
##      White           1084 (77.6)    891 (72.4)
##   AgeDecade (%)                                  0.029
##      10-19             24 ( 1.7)     27 ( 2.2)
##      20-29            322 (23.0)    323 (26.2)
##      30-39            346 (24.8)    245 (19.9)
##      40-49            367 (26.3)    323 (26.2)
##      50-59            338 (24.2)    313 (25.4)
```

Data Visualization

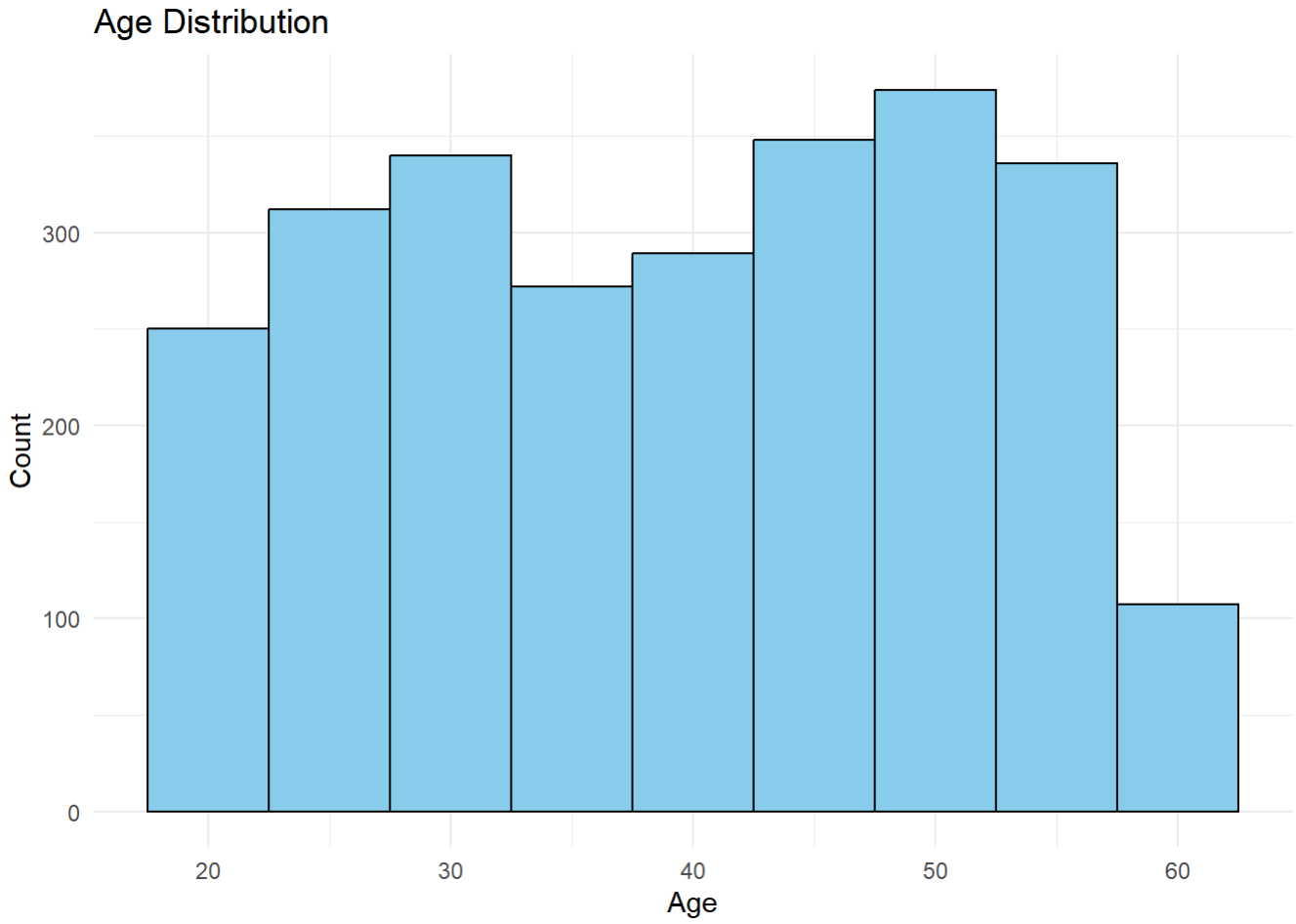
Stacked bar plot of depression severity between Smokers and Non-smokers

```
# Convert Depressed to factor with ordered levels (None, Several, Most)
data_clean$Depressed <- factor(data_clean$Depressed, levels = c(0, 1, 2), labels = c("None",
"Several", "Most"))
ggplot(data_clean, aes(x = RegularMarij, fill = Depressed)) +
  geom_bar(position = "fill") +
  labs(title = "Depression Severity among Marijuana Smokers and Non-Smokers",
       x = "Marijuana Use (Smoker vs. Non-Smoker)",
       y = "Proportion",
       fill = "Depression Severity") +
  scale_y_continuous(labels = scales::percent_format()) +
  theme_minimal() +
  scale_fill_manual(values = c("lightgreen", "orange", "red"))
```



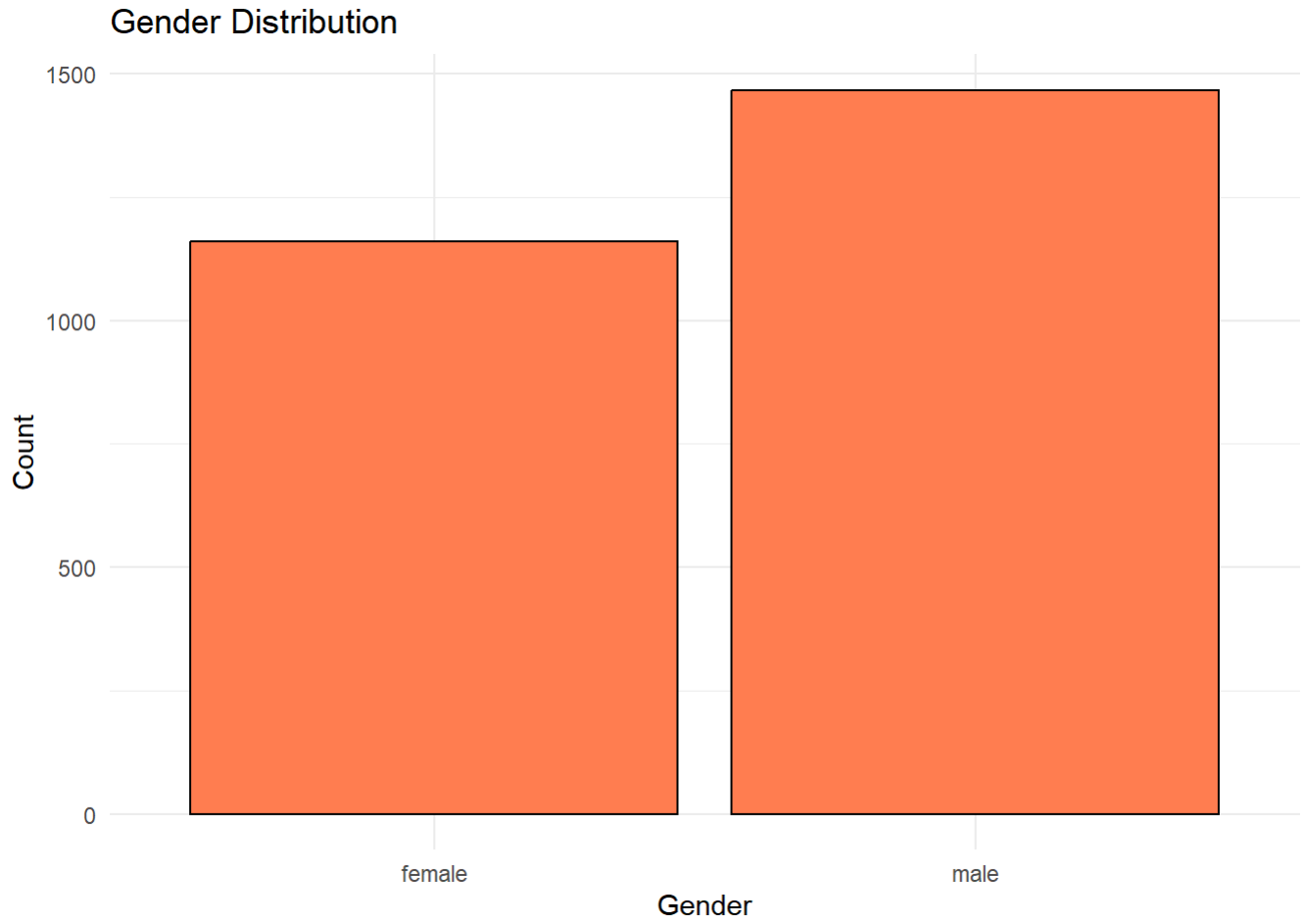
Histogram for Age distribution

```
ggplot(data_clean, aes(x = Age)) +
  geom_histogram(binwidth = 5, fill = "skyblue", color = "black") +
  labs(title = "Age Distribution", x = "Age", y = "Count") +
  theme_minimal()
```



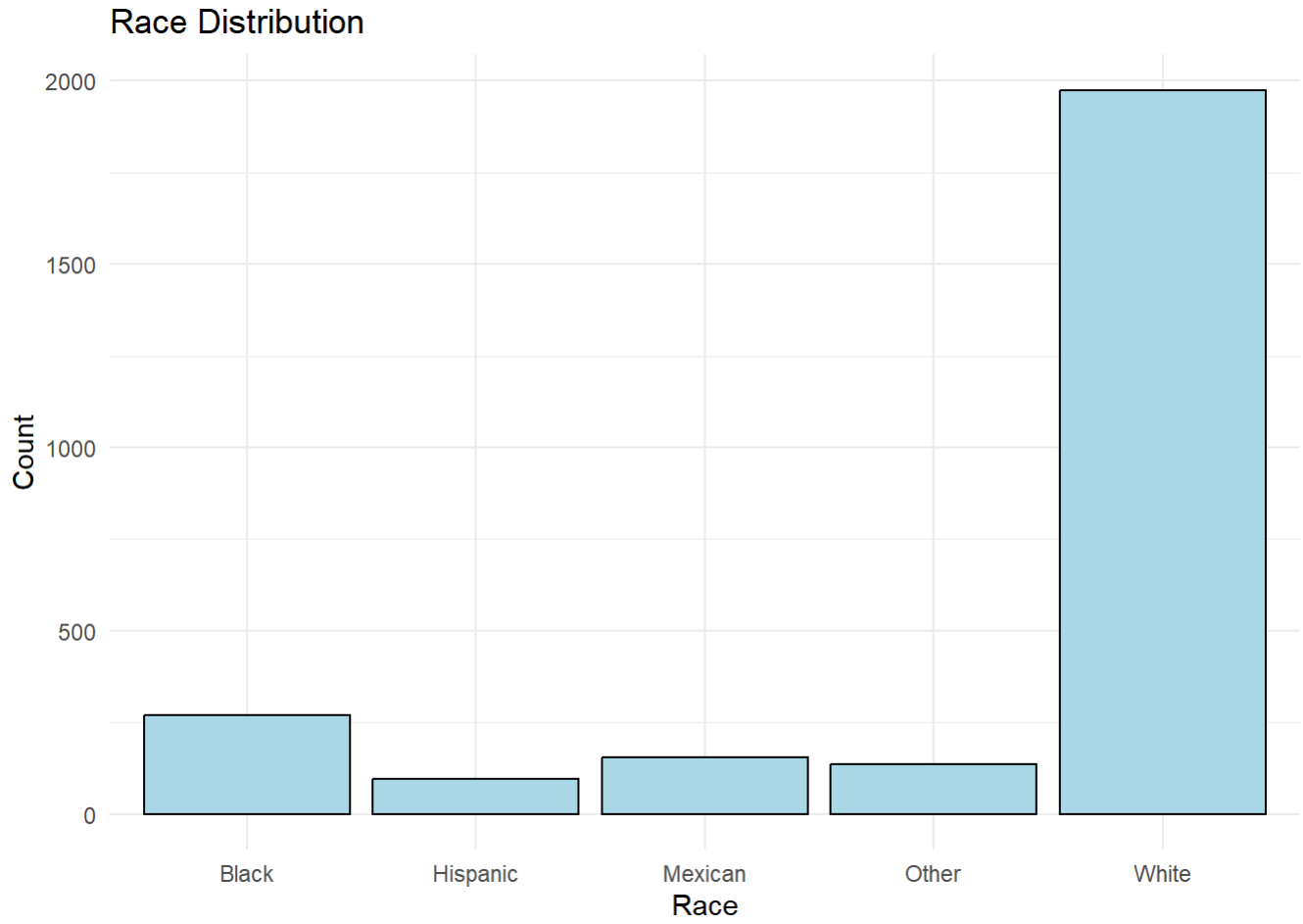
Bar Plot for Gender

```
ggplot(data_clean, aes(x = Gender)) +
  geom_bar(fill = "coral", color = "black") +
  labs(title = "Gender Distribution", x = "Gender", y = "Count") +
  theme_minimal()
```



Bar Plot for Race

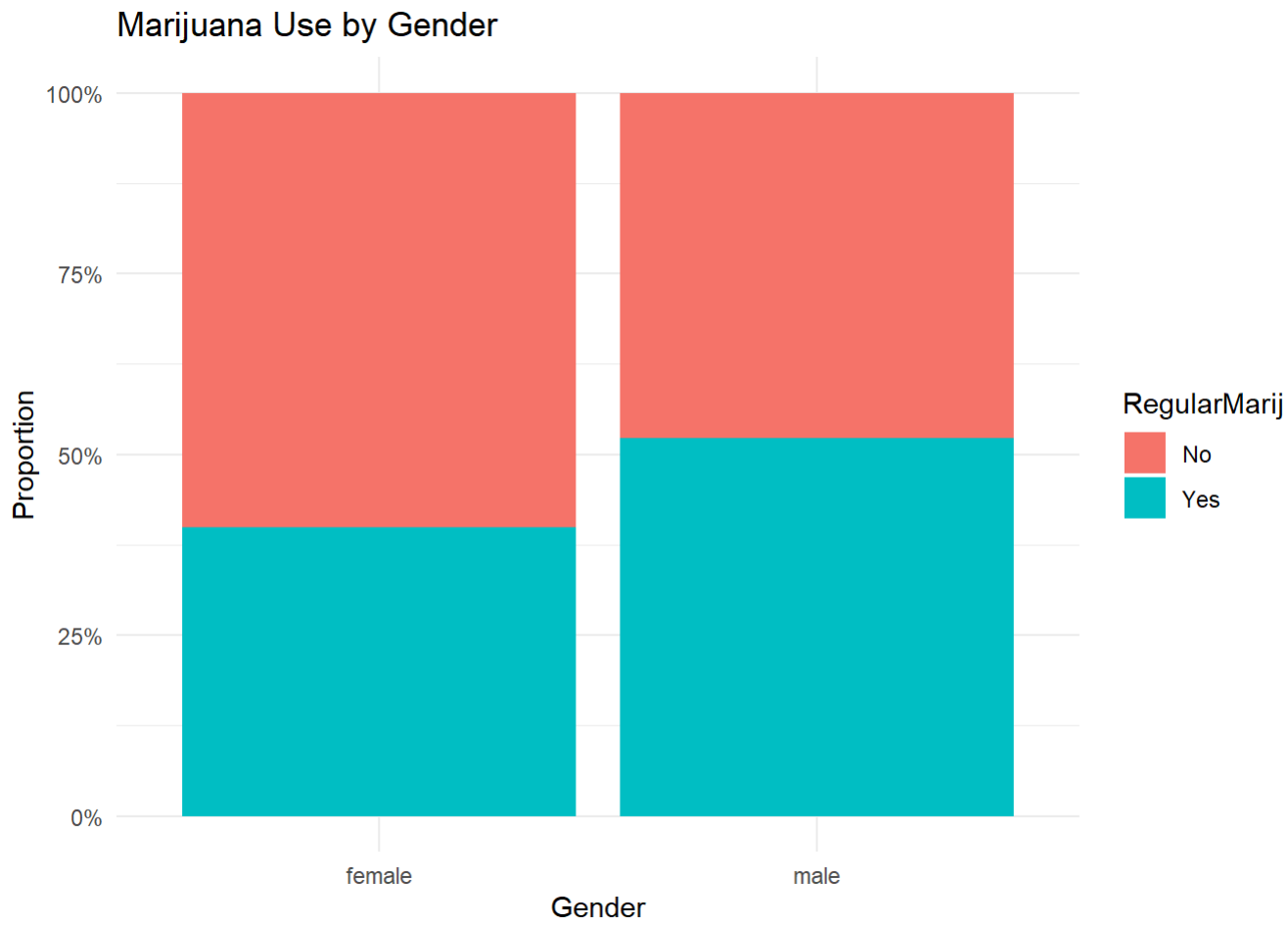
```
ggplot(data_clean, aes(x = Race1)) +
  geom_bar(fill = "lightblue", color = "black") +
  labs(title = "Race Distribution", x = "Race", y = "Count") +
  theme_minimal()
```



Marijuana use by Demogrphics

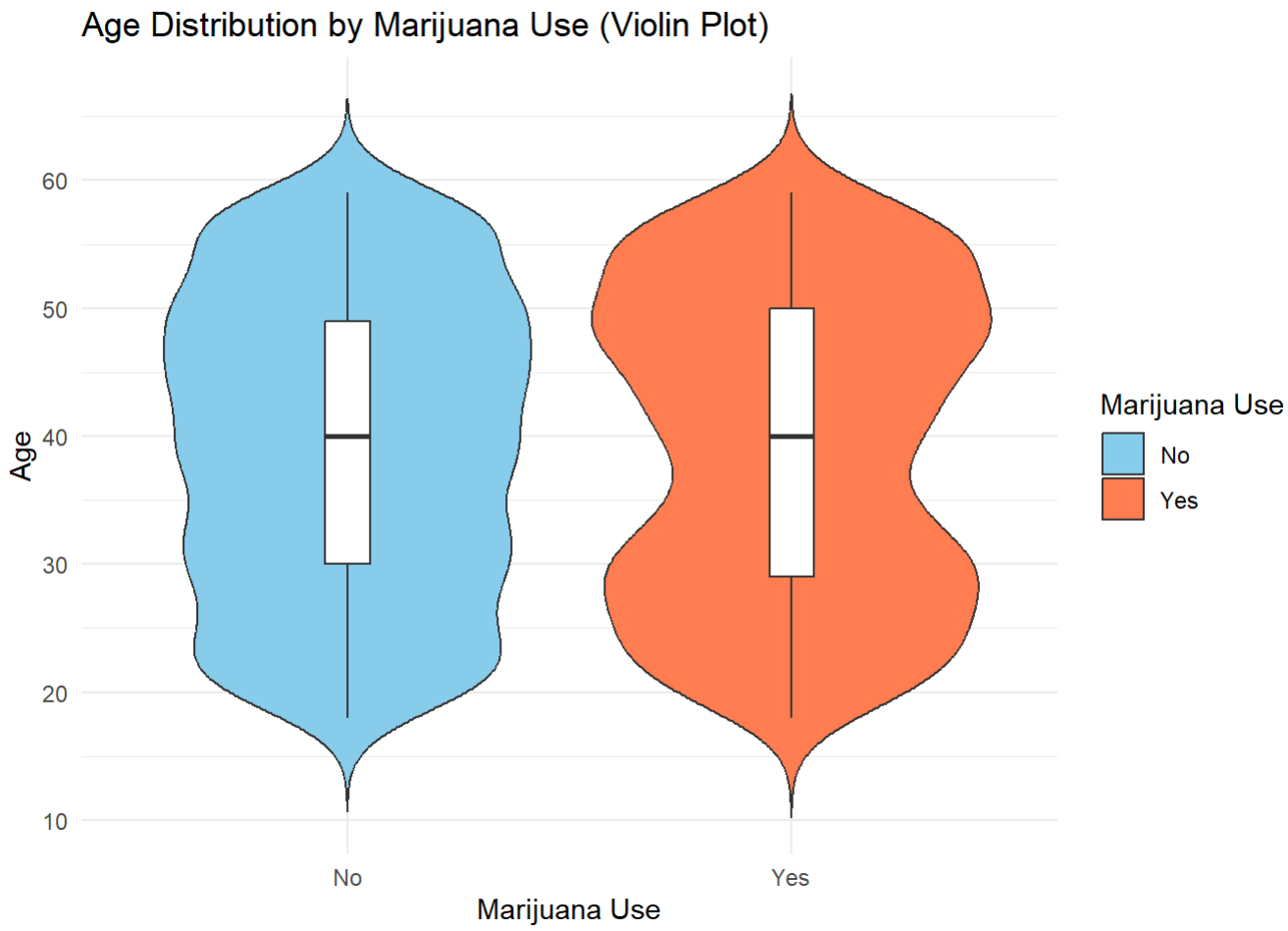
By Gender

```
ggplot(data_clean, aes(x = Gender, fill = RegularMarij)) +
  geom_bar(position = "fill") +
  labs(title = "Marijuana Use by Gender", x = "Gender", y = "Proportion") +
  scale_y_continuous(labels = scales::percent_format()) +
  theme_minimal()
```



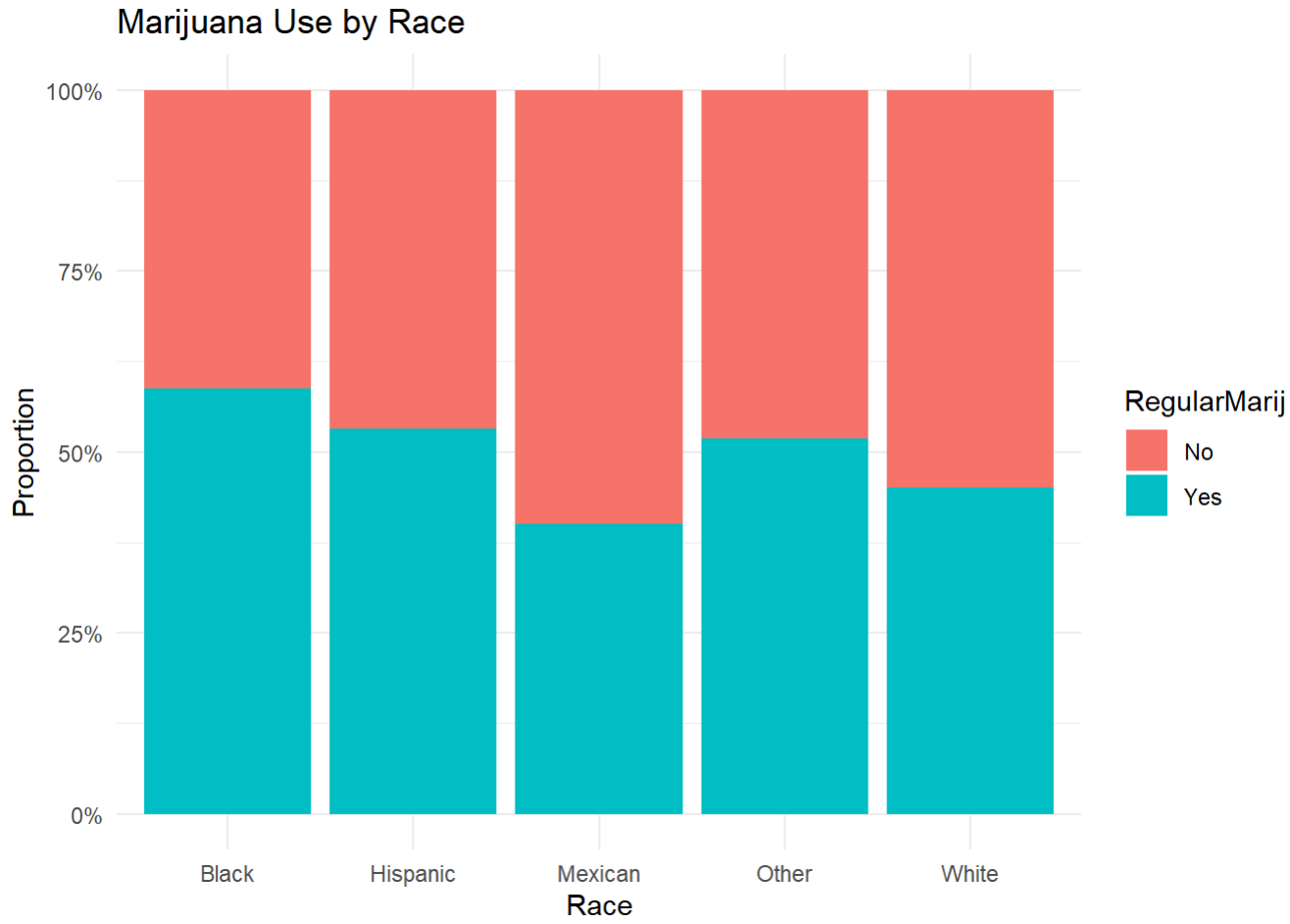
By age distribution

```
ggplot(data_clean, aes(x = RegularMarij, y = Age, fill = RegularMarij)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, fill = "white") +
  labs(title = "Age Distribution by Marijuana Use (Violin Plot)", x = "Marijuana Use", y = "A
ge") +
  theme_minimal() +
  scale_fill_manual(values = c("skyblue", "coral"), name = "Marijuana Use")
```



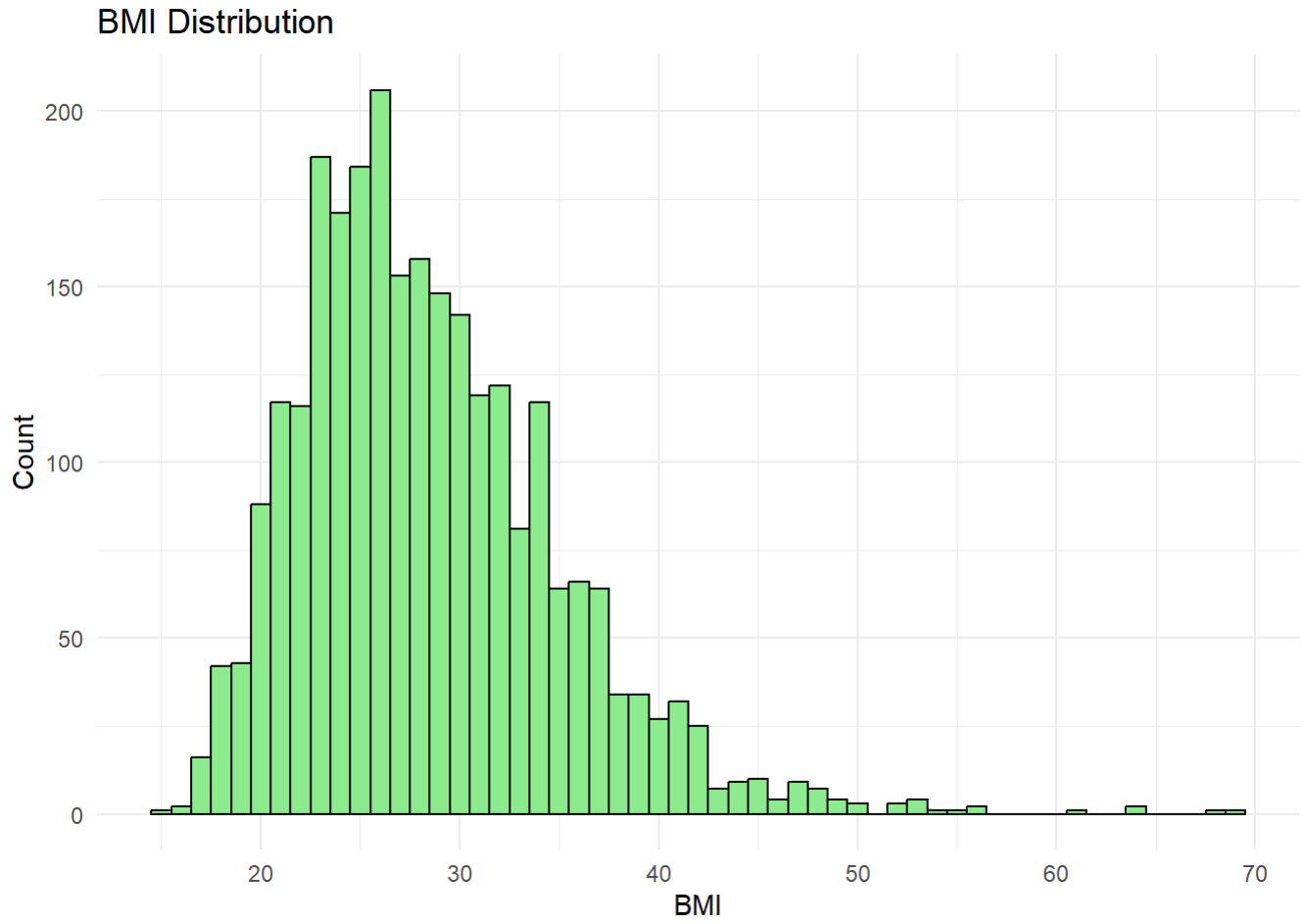
By race

```
ggplot(data_clean, aes(x = Race1, fill = RegularMarij)) +
  geom_bar(position = "fill") +
  labs(title = "Marijuana Use by Race", x = "Race", y = "Proportion") +
  scale_y_continuous(labels = scales::percent_format()) +
  theme_minimal()
```

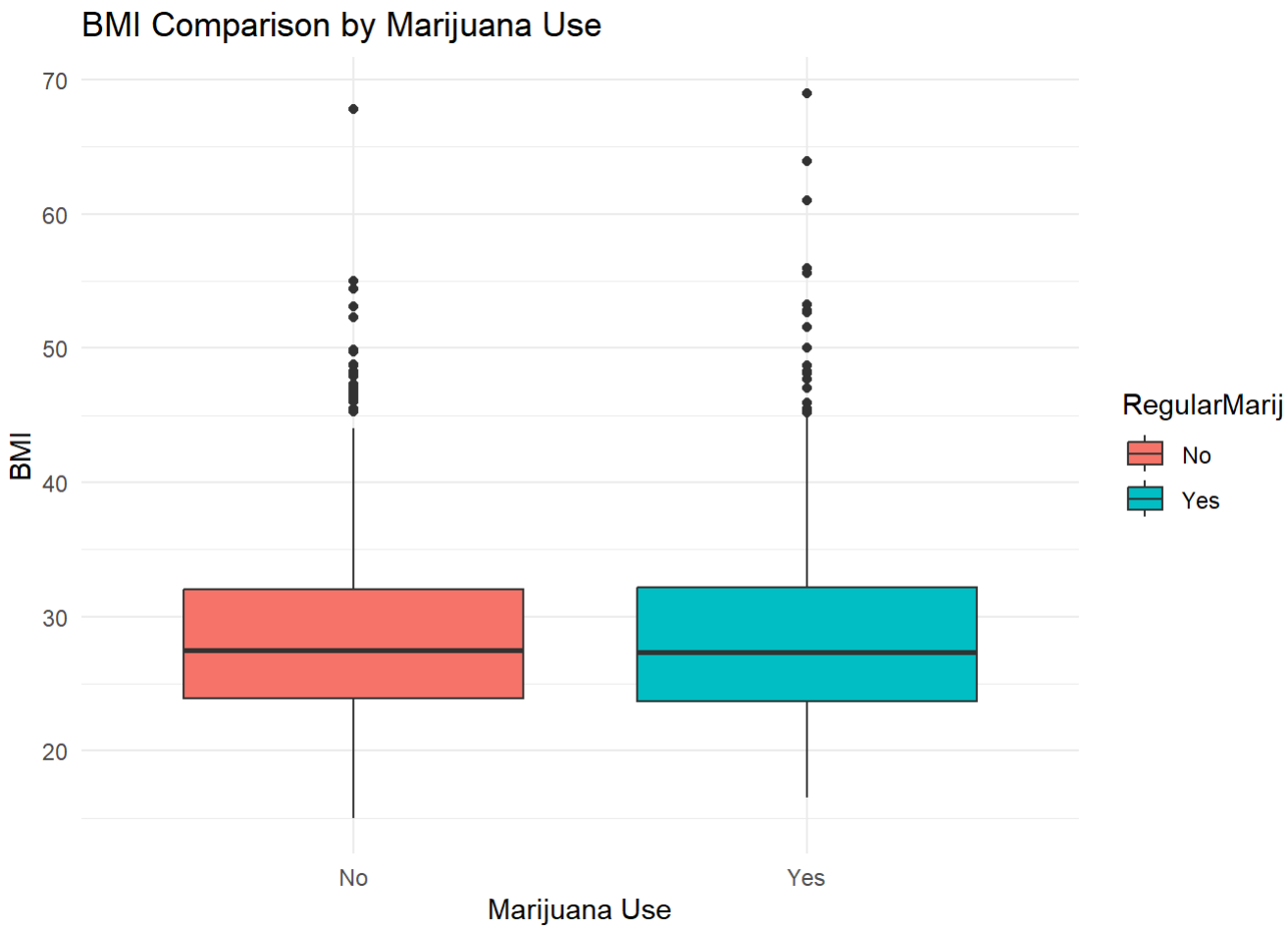


Comparison Visualtisation

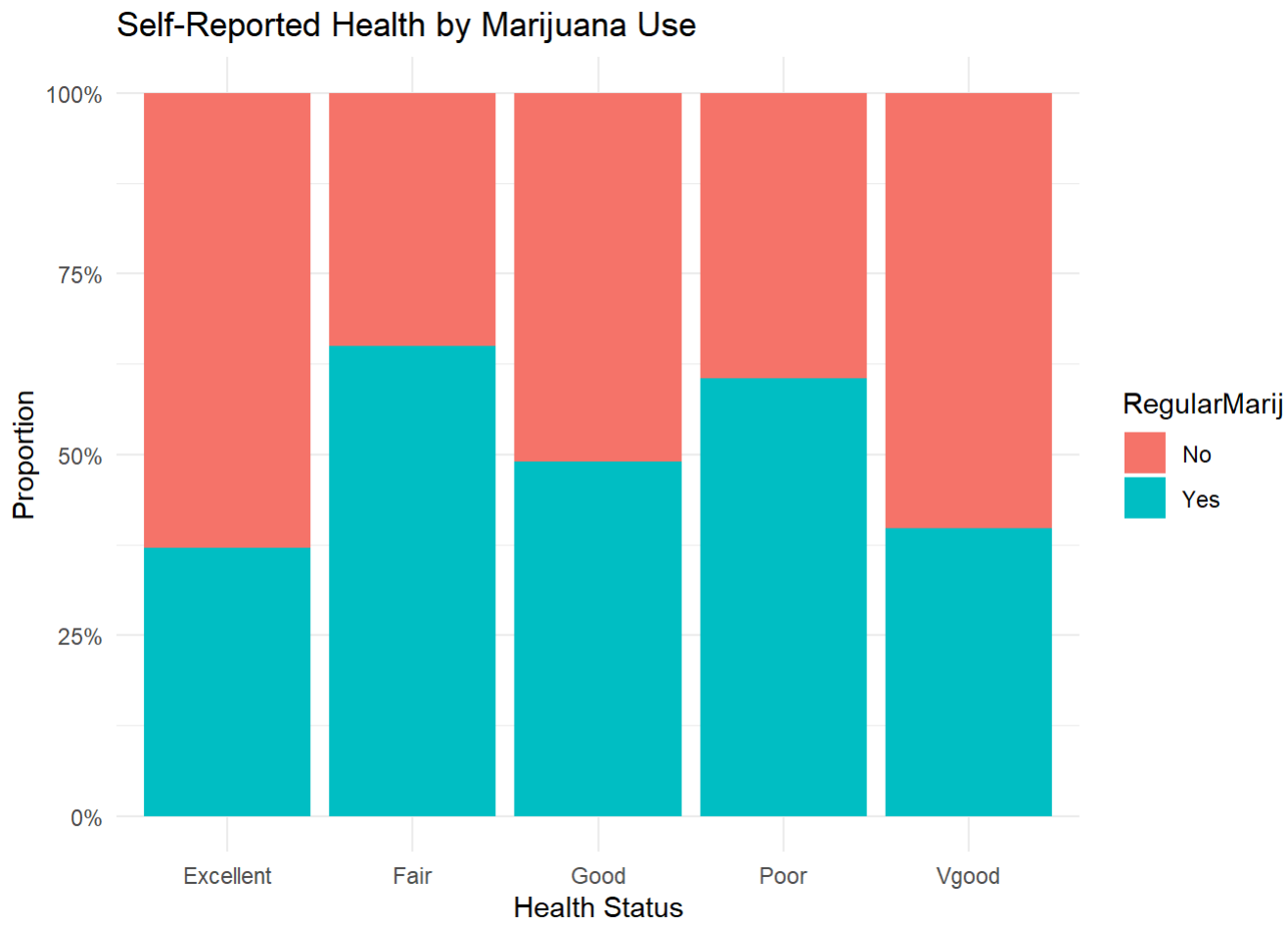
```
#BMI
ggplot(data_clean, aes(x = BMI)) +
  geom_histogram(binwidth = 1, fill = "lightgreen", color = "black") +
  labs(title = "BMI Distribution", x = "BMI", y = "Count") +
  theme_minimal()
```




```
#BMI distribution between smokers and non smokers
ggplot(data_clean, aes(x = RegularMarij, y = BMI, fill = RegularMarij)) +
  geom_boxplot() +
  labs(title = "BMI Comparison by Marijuana Use", x = "Marijuana Use", y = "BMI") +
  theme_minimal()
```




```
# General Health distribution comparison with Marijuana use
ggplot(data_clean, aes(x = HealthGen, fill = RegularMarij)) +
  geom_bar(position = "fill") +
  labs(title = "Self-Reported Health by Marijuana Use", x = "Health Status", y = "Proportio
n") +
  scale_y_continuous(labels = scales::percent_format()) +
  theme_minimal()
```



Analysis

Maarijuana use and Physical Health

Using T-test to check if there is a difference in BMI between Marijuana smokers and non-smokers

```
#T test for difference between BMI in smokers and non-smokers
test1 <- t.test(BMI ~ RegularMarij, data = data_clean)
print(t)
```

```
## function (x)
## UseMethod("t")
## <bytecode: 0x0000026590922660>
## <environment: namespace:base>
```

Interpretation:

1. The t-value of 0.21062 indicates a very small difference between the means of BMI for marijuana smokers and non-smokers.
2. The p-value of 0.8332 is much greater than the typical significance threshold of 0.05. This means there is no statistically significant difference between the mean BMI of marijuana smokers and non-smokers.
3. The 95% confidence interval (-0.445, 0.552) includes zero, which further supports the conclusion that the difference in BMI between the two groups is not statistically significant.
4. Sample estimates show that the average BMI for both groups is almost identical (28.5 for non-smokers and 28.45 for smokers).

Marijuana use and Mental Heatlh

To analyse our hypthesis regarding to mental health, we use Depression variable in the data set as a proxy for mental health status. Depression “Yes” or “NO” is engineered to indicate clear differnce between depressed and non depressed individuals

Chi-square test “Marijuana use” and “Depression” using a 2x2 contingency table

```
contingency_table <- table(data_clean$RegularMarij, data_clean$Depressed)
dimnames(contingency_table) <- list("Marijuana Use" = c("No", "Yes"),
                                    "Depression Level" = c("None", "Several", "Most"))
print(contingency_table)
```

```
##              Depression Level
## Marijuana Use None Several Most
##           No  1096     226   75
##           Yes  882     244  105
```

```
chisq.test(table(data_clean$RegularMarij, data_clean$Depressed))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(data_clean$RegularMarij, data_clean$Depressed)
## X-squared = 18.43, df = 2, p-value = 9.953e-05
```

Interpretation

The **Chi-Square test** result indicates a significant association between marijuana use and depression, since the p-value is much smaller than typical significance thresholds (e.g., 0.05). Therefore, we can conclude that marijuana use is associated with depression in your sample, but it does not indicate the direction of the relationship

Logistic Regression

To model affect of marijuana consumption and depression keeping age and gender like factors as control variables

```
data_clean$DepressedBinary <- ifelse(data_clean$Depressed %in% c(0, 1), 0, 1)
data_clean$Age <- scale(data_clean$Age)
data_clean$HHIncomeMid <- scale(data_clean$HHIncomeMid)
library(logistf)
```

```
## Warning: package 'logistf' was built under R version 4.4.2
```

```
# Running the logistic regression with increased max iterations
full_model <- logistf(formula = DepressedBinary ~ RegularMarij + Age + Gender,
                      data = data_clean, control = logistf.control(maxit = 100))

summary(full_model)
```

```
## logistf(formula = DepressedBinary ~ RegularMarij + Age + Gender,
##     data = data_clean, control = logistf.control(maxit = 100))
##
## Model fitted by Penalized ML
## Coefficients:
##                         coef  se(coef) lower 0.95 upper 0.95       Chisq
## (Intercept)      7.14410009 1.1887535   5.219474  12.204580         Inf
## RegularMarijYes -0.17431458 1.4254527  -5.439082   5.088137 0.007232601
## Age              0.05345368 0.7070514  -6.671946   7.199680 0.001820301
## Gendermale       0.23286661 1.4256795  -5.029929   5.501989 0.012875423
##                          p method
## (Intercept)     0.0000000      2
## RegularMarijYes 0.9322259      2
## Age             0.9659686      2
## Gendermale      0.9096580      2
##
## Method: 1-Wald, 2-Profile penalized log-likelihood, 3-None
##
## Likelihood ratio test=0.01887541 on 3 df, p=0.9993142, n=2628
## Wald test = 102.8024 on 3 df, p = 0
```

Interpretation

Coefficients Interpretation: (Intercept): The coefficient for the intercept is 7.144. This is the log-odds of depression when all other variables (RegularMarij, Age, Gender) are at their reference values (i.e., no marijuana use, Age = 0, Gender = female).

1. **RegularMarijYes:** The coefficient for marijuana use is -0.174, but the p-value is 0.932, which is not statistically significant. This suggests that there is no clear evidence to support that marijuana use is associated with a higher or lower likelihood of depression after controlling for age and gender.
2. **Age:** The coefficient for age is 0.053, but the p-value is 0.966, which is also not statistically significant. This suggests that age does not have a meaningful association with depression after controlling for marijuana use and gender.
3. **Gendermale:** The coefficient for being male is 0.233, but the p-value is 0.910, which is not statistically significant. This suggests that gender does not have a significant association with depression after accounting for marijuana use and age.
4. **Likelihood Ratio Test (p-value = 0.999):** The p-value from the likelihood ratio test is extremely high, suggesting that the model as a whole does not significantly improve the fit compared to a null model (a model with no predictors). This reinforces the interpretation that marijuana use, age, and gender may not be strong predictors of depression in this dataset.

Conclusion

Based on the analysis, the results suggest that marijuana use does not have a significant effect on depression when controlling for other factors such as age, gender, and household income. Specifically, the logistic regression models did not show a statistically significant association between marijuana consumption and the likelihood of being depressed. The p-values for marijuana use (RegularMarij) in both the univariate and multivariate models were not significant, indicating that marijuana use does not have a strong effect on depression after adjusting for confounders.

Key Findings:

1. **Chi-Square Test:** The chi-square test for the association between marijuana use and depression revealed no significant relationship (p-value = 0.9496), indicating that marijuana consumption and depression prevalence were independent.
2. **Logistic Regression:** The coefficients for marijuana use in the logistic regression models were not significant (p-value > 0.05), suggesting that marijuana use does not significantly predict depression status after controlling for other variables such as age, gender, and household income.
3. **Other Factors:** Age and gender were included as potential confounders in the model. However, none of these variables showed strong evidence of affecting the relationship between marijuana use and depression.

Future Scope:

1. **Longitudanal Data:** A longitudinal study following individuals over a longer period could provide more robust insights into how marijuana consumption affects mental and physical health over time. You could analyze trends, causality, and whether the effects of marijuana change over the years, especially in different age groups.
2. **Different Types of marijuana use:** It would be insightful to categorize marijuana use more specifically (e.g., frequency of use, type of marijuana, method of consumption). Analyzing the different patterns of use (e.g., medical vs recreational, smoking vs edibles) could reveal more nuanced effects on health.
3. **Examine other mental health conditions:** Exploring marijuana’s effects on other mental health conditions (e.g., anxiety, schizophrenia, PTSD) would offer a broader understanding of its potential therapeutic or detrimental effects. This can help in formulating policy recommendations for medical marijuana use.