Epi 3 Data Analysis

Qiheng(Michael) Yan

2023-11-21

```r
library(tidyverse)
library(stats)
library(ggplot2)
library(expss)
set.seed(1028)
```

Epi 3 Data Analysis

Import data

```r
# Import the dataset
complete_data<-readRDS("combined_data.rds")
subset_1168_data<-readRDS("subset_1168.rds")
#write.csv(complete_data, "complete_data.csv", row.names=FALSE)
#write_labelled_csv(complete_data, "complete_data_labeled.csv", row.names=FALSE)
#write.csv(subset_1168_data, "subset_1168_data.csv", row.names=FALSE)
#write_labelled_csv(subset_1168_data, "subset_1168_data_labeled.csv", row.names=FALSE)
```

This analysis follows the inclusion/exclusion criteria and uses a subset of 1168 subjects (using subset_1168_data)

Bivariate Analysis

In bivariate analysis, we'll use Chi-squared tests for categorical variables like health insurance status, race/ethnicity, gender, and a T-test for continuous variables like age.

```r
# Chi-squared test for medication adherence and family income
chisq_income <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$income_cat))

# Chi-squared test for medication adherence and total insurance status
chisq_insurance <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$ins_classif))
```

```
## Warning in chisq.test(table(subset_1168_data$adherence,
## subset_1168_data$ins_classif)): Chi-squared approximation may be incorrect
```

```r
# Chi-squared test for medication adherence and race
chisq_race <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$race_6cat))

# Chi-squared test for medication adherence and sex
chisq_gender <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$sex))

# Chi-squared test for medication adherence and education level
chisq_edu <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$educ_level))

# Chi-squared test for medication adherence and marital status
chisq_marital <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$marital_status))

# Chi-squared test for medication adherence and alcohol use
chisq_alcohol <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$alc_heavy))

# Chi-squared test for medication adherence and tobacco use
chisq_tobacco <- chisq.test(table(subset_1168_data$adherence, subset_1168_data$smoking_hx))

# T-test for medication adherence and age
t_test_age <- t.test(subset_1168_data$age ~ subset_1168_data$adherence)

# Print the results
print(chisq_income)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(subset_1168_data$adherence, subset_1168_data$income_cat)
## X-squared = 7.6725, df = 2, p-value = 0.02157
```

```r
print(chisq_insurance)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(subset_1168_data$adherence, subset_1168_data$ins_classif)
## X-squared = 32.62, df = 8, p-value = 7.209e-05
```

```r
print(chisq_race)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(subset_1168_data$adherence, subset_1168_data$race_6cat)
## X-squared = 9.8825, df = 5, p-value = 0.07863
```

```r
print(chisq_gender)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(subset_1168_data$adherence, subset_1168_data$sex)
## X-squared = 4.1148, df = 1, p-value = 0.04251
```

```
print(chisq_edu)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(subset_1168_data$adherence, subset_1168_data$educ_level)
## X-squared = 2.2859, df = 4, p-value = 0.6833
```

```
print(chisq_marital)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(subset_1168_data$adherence, subset_1168_data$marital_status)
## X-squared = 14.11, df = 2, p-value = 0.0008633
```

```
print(chisq_alcohol)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(subset_1168_data$adherence, subset_1168_data$alc_heavy)
## X-squared = 3.7807, df = 1, p-value = 0.05185
```

```
print(chisq_tobacco)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(subset_1168_data$adherence, subset_1168_data$smoking_hx)
## X-squared = 14.247, df = 2, p-value = 0.0008058
```

```
print(t_test_age)
```

```
##
##  Welch Two Sample t-test
##
## data:  subset_1168_data$age by subset_1168_data$adherence
## t = -8.921, df = 692.18, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group FALSE and group TRUE is not equal to (
## 95 percent confidence interval:
##  -8.734877 -5.583584
## sample estimates:
## mean in group FALSE  mean in group TRUE
##           57.66085            64.82008
```

The results from the bivariate analyses provide valuable insights into the relationships between adherence to prescribed cholesterol medication and various factors such as family income, health insurance status, race/ethnicity, gender, marital status, alcohol use, tobacco use, and age.

1. **Family Income vs. Adherence**:

Chi-squared test result:
$$X^2 = 7.6725, df = 2, p-value = 0.02157$$

Interpretation: There is a statistically significant association between family income and medication adherence. The low p-value suggests that differences in medication adherence across different income categories are not likely due to chance.

2. **Health Insurance Status vs. Adherence**:

Chi-squared test result:

$$X^2 = 32.62, df = 8, p-value = 7.209 \times 10^{-5}$$

Interpretation: There is a strong statistically significant association between insurance status and medication adherence. The very low p-value indicates that the observed differences in adherence across various insurance classifications are unlikely to be due to random variation.

3. **Race/Ethnicity vs. Adherence**:

Chi-squared test result:
$$X^2 = 9.8825, df = 5, p-value = 0.07863$$

Interpretation: The association between race and medication adherence is not statistically significant at the conventional alpha level of 0.05. This suggests that differences in adherence across different racial categories may be due to chance.

4. **Gender vs. Adherence**:

Chi-squared test result:
$$X^2 = 4.1148, df = 1, p-value = 0.04251$$

Interpretation: There is a statistically significant association between gender and medication adherence, with the p-value indicating that these differences are unlikely to be due to random chance.

5. **Education Level vs. Adherence**:

Chi-squared test result:
$$X^2 = 2.2859, df = 4, p-value = 0.6833$$

Interpretation: The p-value of 0.6833 is well above the conventional alpha level of 0.05, which suggests that there is no statistically significant association between education level and adherence to cholesterol medication in this sample.

6. **Marital Status vs. Adherence**:

Chi-squared test result:
$$X^2 = 14.11, df = 2, p-value = 0.0008633$$

Interpretation: This suggests that marital status is an important factor in adherence.

7. **Alcohol Use vs. Adherence**:

Chi-squared test result:
$$X^2 = 3.7807, df = 1, p-value = 0.05185$$

Interpretation: The association between heavy alcohol use and medication adherence is on the borderline of statistical significance, indicating a potential relationship that may require further exploration.

8. **Tobacco use vs. Adherence**:

Chi-squared test result:
$$X^2 = 14.247, df = 2, p-value = 0.0008058$$

Interpretation: The association between heavy alcohol use and medication adherence is on the borderline of statistical significance, indicating a potential relationship that may require further exploration.

9. **Age vs. Adherence**:

T-test result:
$$t = -8.921, df = 692.18, p-value = 2.2 \times 10^{-16}$$

Interpretation: There is a significant association between tobacco use and medication adherence, suggesting that smoking status is an important factor.

Note on the Warning: The warning regarding the Chi-squared approximation may be due to small expected frequencies in some cells of the contingency tables. This is common when some categories have a low number of observations. In such cases, alternative tests like Fisher's Exact Test for small sample sizes might be more appropriate, especially for the race and gender analyses.

Bivariate Analysis Interpretation

The bivariate analysis indicates several noteworthy relationships between demographic, socio-economic, and behavioral factors and medication adherence. Income level, insurance status, marital status, and tobacco use all show statistically significant correlations with adherence to cholesterol medication, suggesting these factors may play a critical role in whether individuals follow their prescribed medication regimen. Additionally, age remains a strong factor, with older individuals more likely to adhere to their medication. Although the associations with race/ethnicity and alcohol use did not reach statistical significance, they may still be relevant in the broader context of medication adherence and are worth exploring further. These bivariate findings lay the groundwork for more comprehensive analyses that can untangle the complex interplay of these variables and inform targeted strategies for improving adherence.

Bivariate Analysis Visualization

```
# Ensure NA is a factor level and label it as "No Response"
subset_1168_data$adherence_factor <- factor(subset_1168_data$adherence, levels = c(FALSE, TRUE))
subset_1168_data$adherence_factor <- addNA(subset_1168_data$adherence_factor)
levels(subset_1168_data$adherence_factor)[is.na(levels(subset_1168_data$adherence_factor))] <- "No Resp

# Define new colors for the bars, including NA
new_colors <- c("TRUE" = "#1b9e77", "FALSE" = "#d95f02", "No Response" = "#4D4D4D")

# Create the plots, making sure to use scale_fill_manual to include NA values
# and set the axis titles correctly after coord_flip()
```
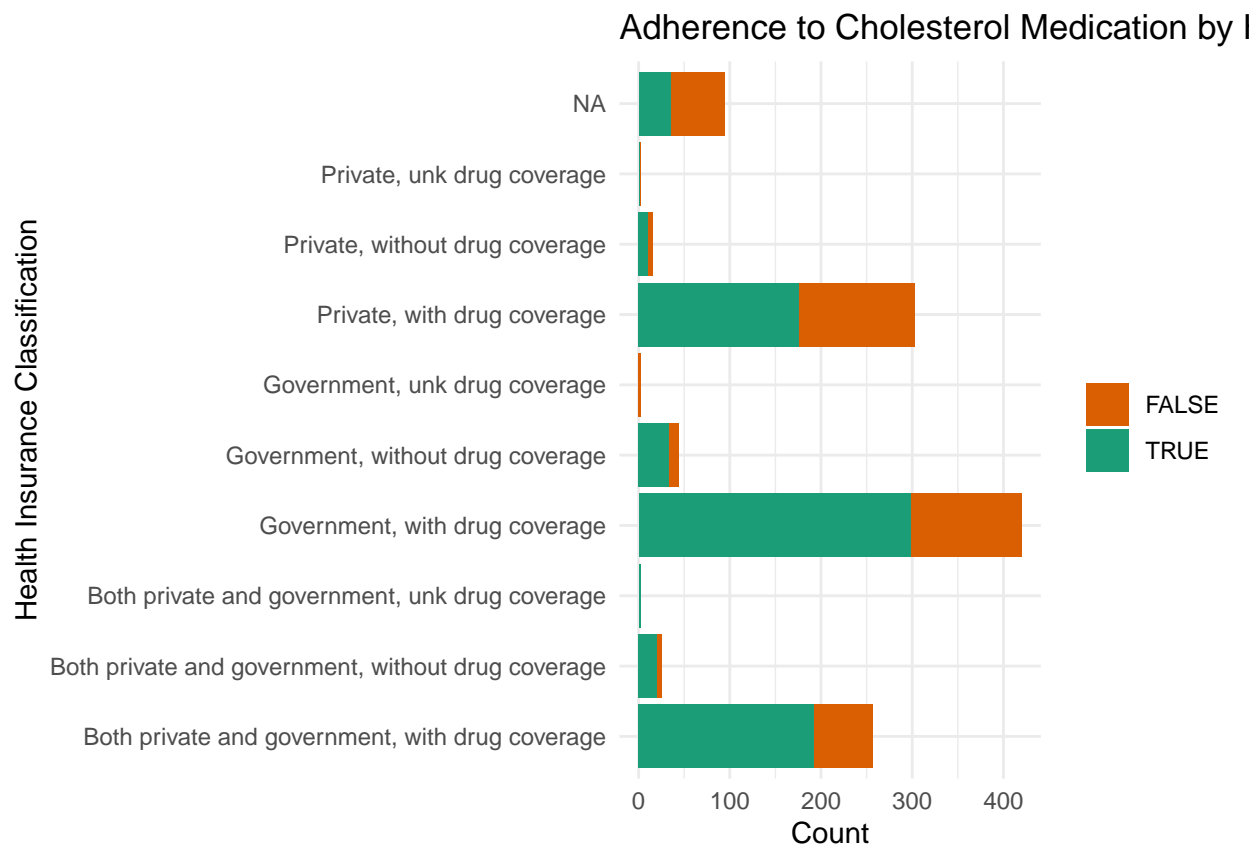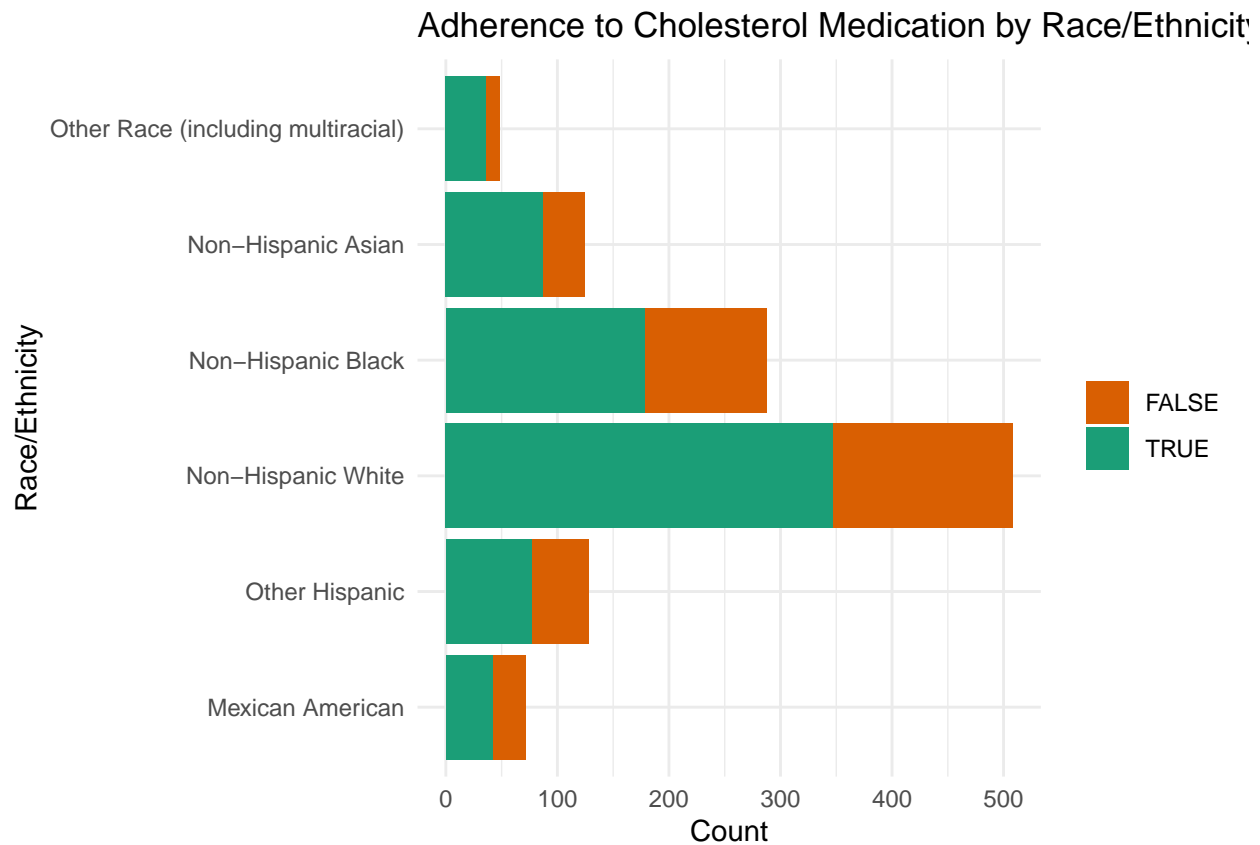
```
# Income Category vs Adherence
ggplot(subset_1168_data, aes(x = income_cat, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Income Category",
       y = "Count",
       x = "Income Category") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```
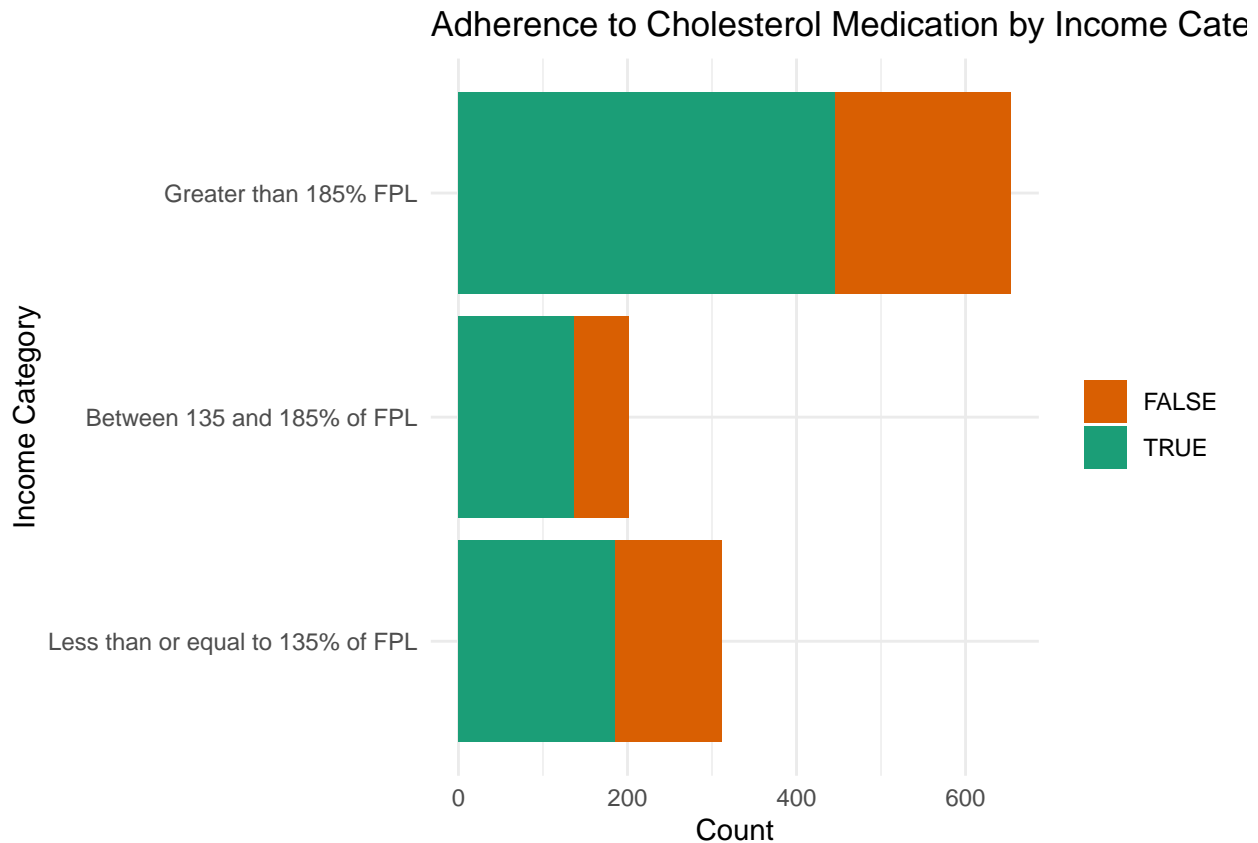


```
# Health Insurance Classification vs Adherence
ggplot(subset_1168_data, aes(x = ins_classif, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Health Insurance Status",
       y = "Count",
       x = "Health Insurance Classification") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```
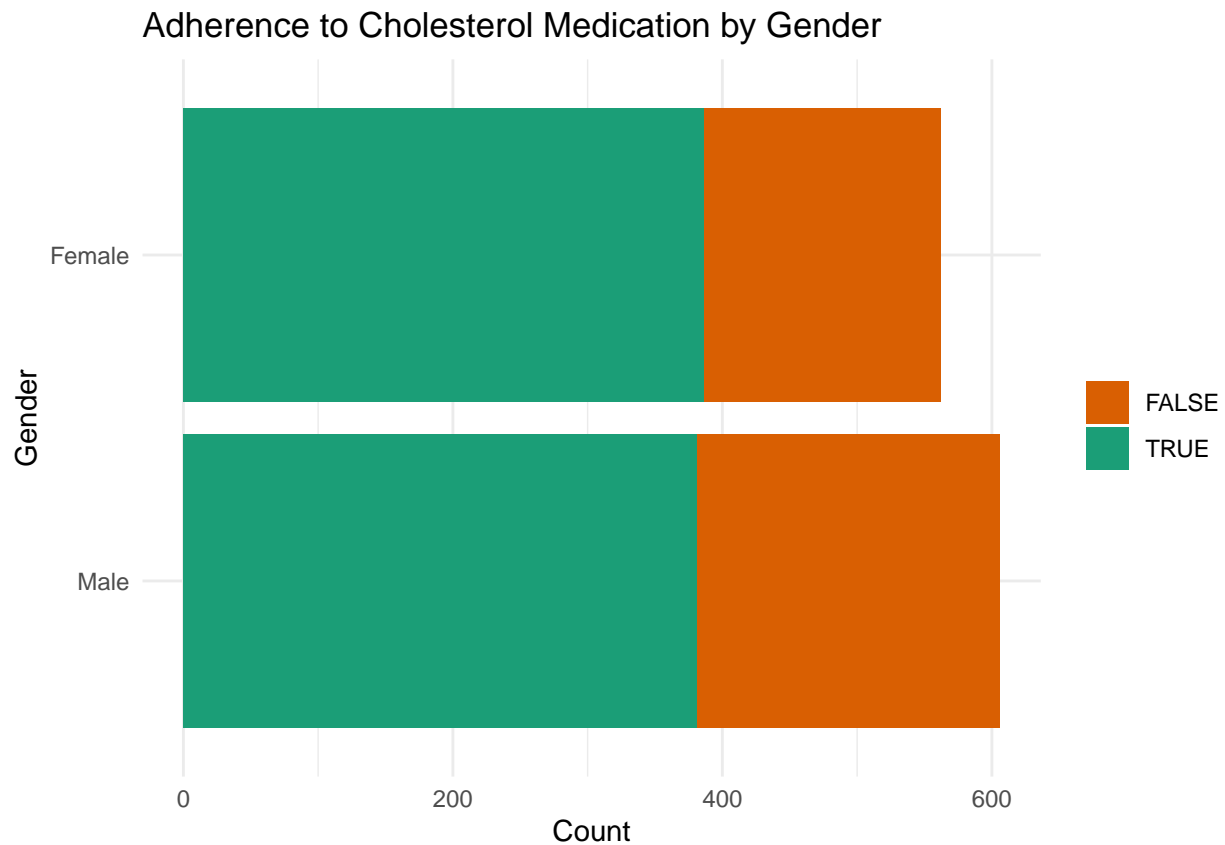
Adherence to Cholesterol Medication by I

```r
# Race/Ethnicity vs Adherence
ggplot(subset_1168_data, aes(x = race_6cat, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Race/Ethnicity",
       y = "Count",
       x = "Race/Ethnicity") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```

Adherence to Cholesterol Medication by Race/Ethnicity
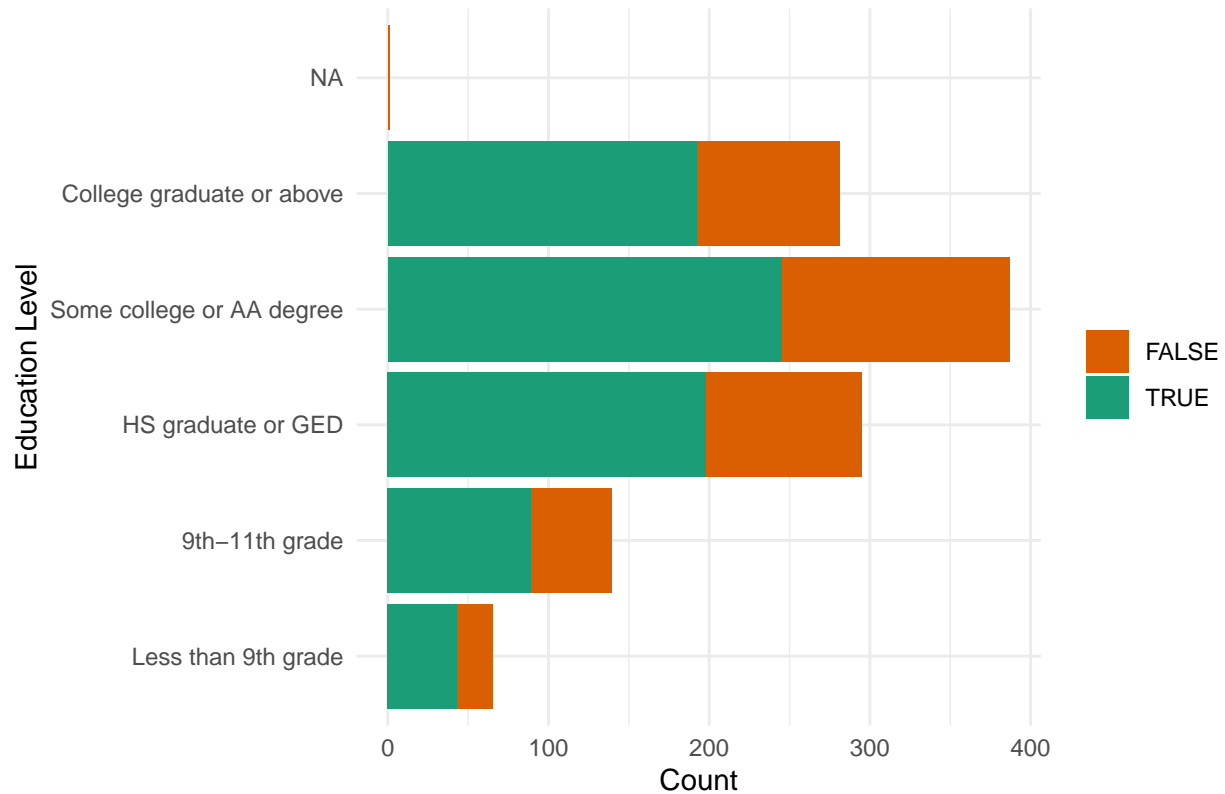


```r
# Gender vs Adherence
ggplot(subset_1168_data, aes(x = sex, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Gender",
       y = "Count",
       x = "Gender") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```

Adherence to Cholesterol Medication by Gender
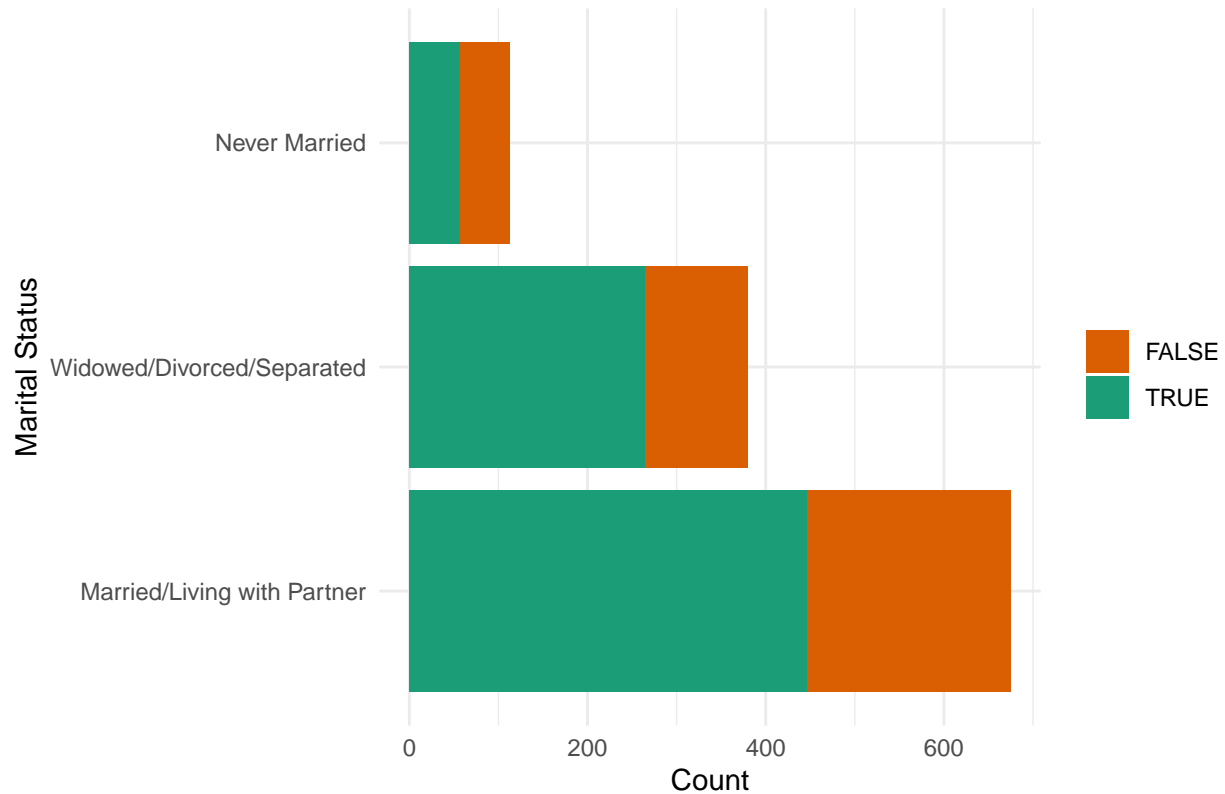


```r
# Education Level vs Adherence
ggplot(subset_1168_data, aes(x = educ_level, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Education Level",
       y = "Count",
       x = "Education Level") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```

Adherence to Cholesterol Medication by Education Level


```r
# Marital Status vs Adherence
ggplot(subset_1168_data, aes(x = marital_status, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Marital Status",
       y = "Count",
       x = "Marital Status") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```

Adherence to Cholesterol Medication by Marital Status

```
# Alcohol Use vs Adherence
ggplot(subset_1168_data, aes(x = alc_heavy, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Alcohol Use",
      y = "Count",
      x = "Alcohol Use") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```

Adherence to Cholesterol Medication by Alcohol Use


```r
# Tobacco Use vs Adherence
ggplot(subset_1168_data, aes(x = smoking_hx, fill = adherence_factor)) +
  geom_bar() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Adherence to Cholesterol Medication by Tobacco Use",
      y = "Count",
      x = "Tobacco Use") +
  theme_minimal() +
  coord_flip() +
  theme(legend.title = element_blank())
```

Adherence to Cholesterol Medication by Tobacco Use


```r
# Age vs Adherence Box Plot
ggplot(subset_1168_data, aes(x = adherence_factor, y = age, fill = adherence_factor)) +
  geom_boxplot() +
  scale_fill_manual(values = new_colors) +
  labs(title = "Age Distribution by Cholesterol Medication Adherence",
       x = "Adherence to Medication",
       y = "Age") +
  theme_minimal() +
  theme(legend.title = element_blank())
```
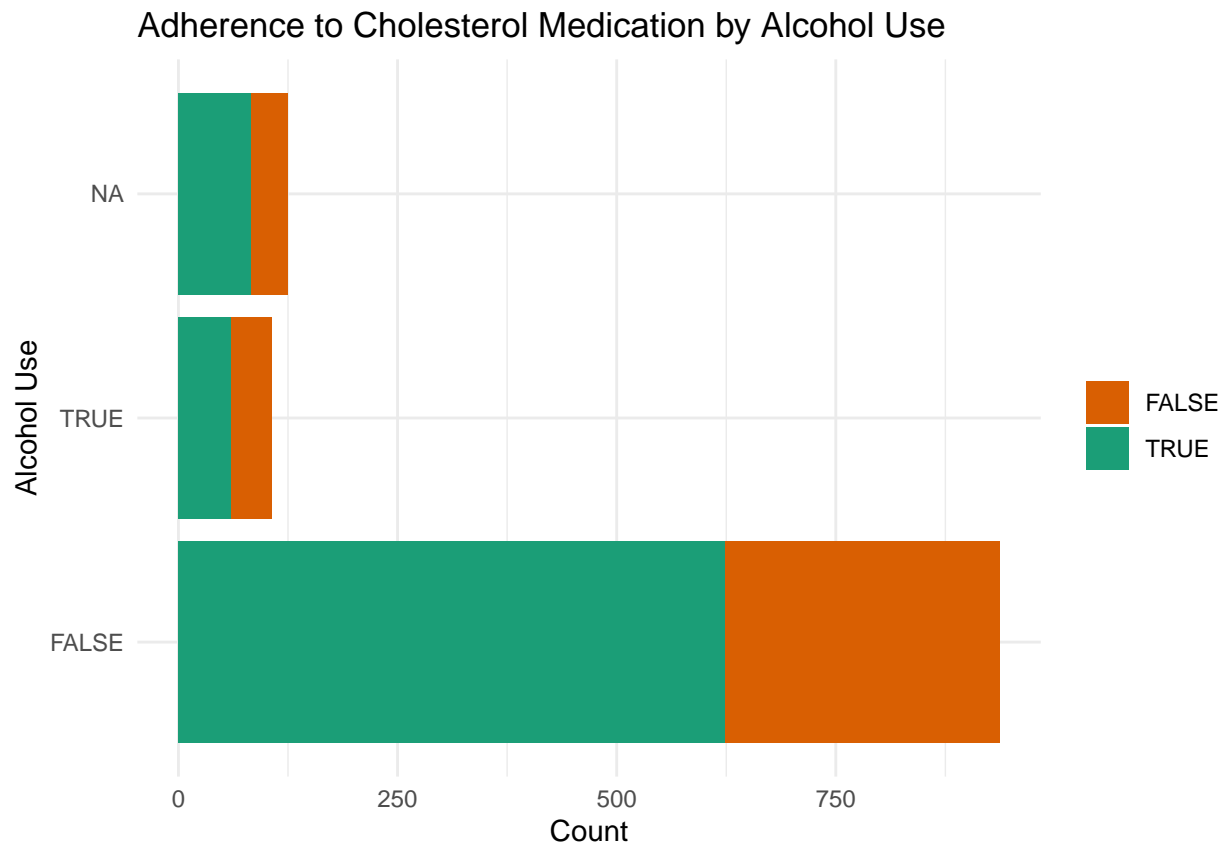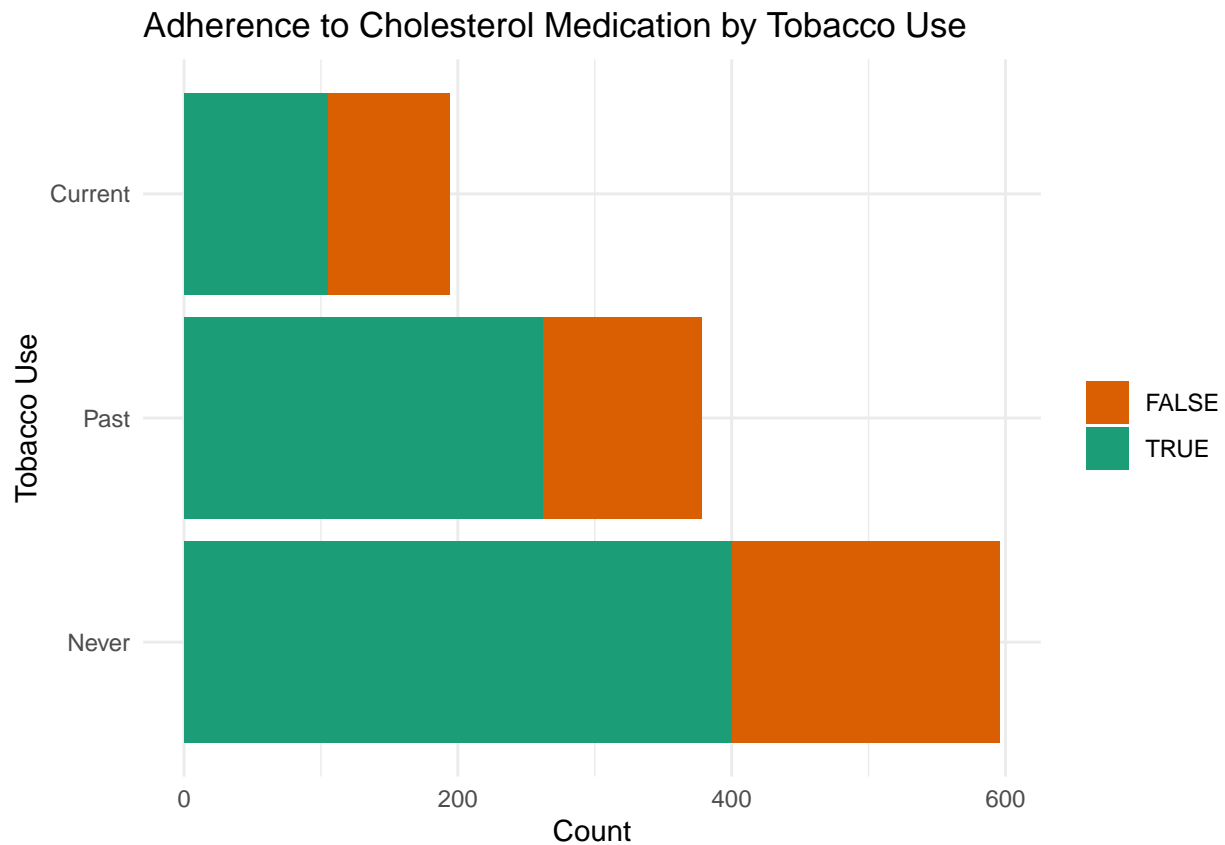
Age Distribution by Cholesterol Medication Adherence
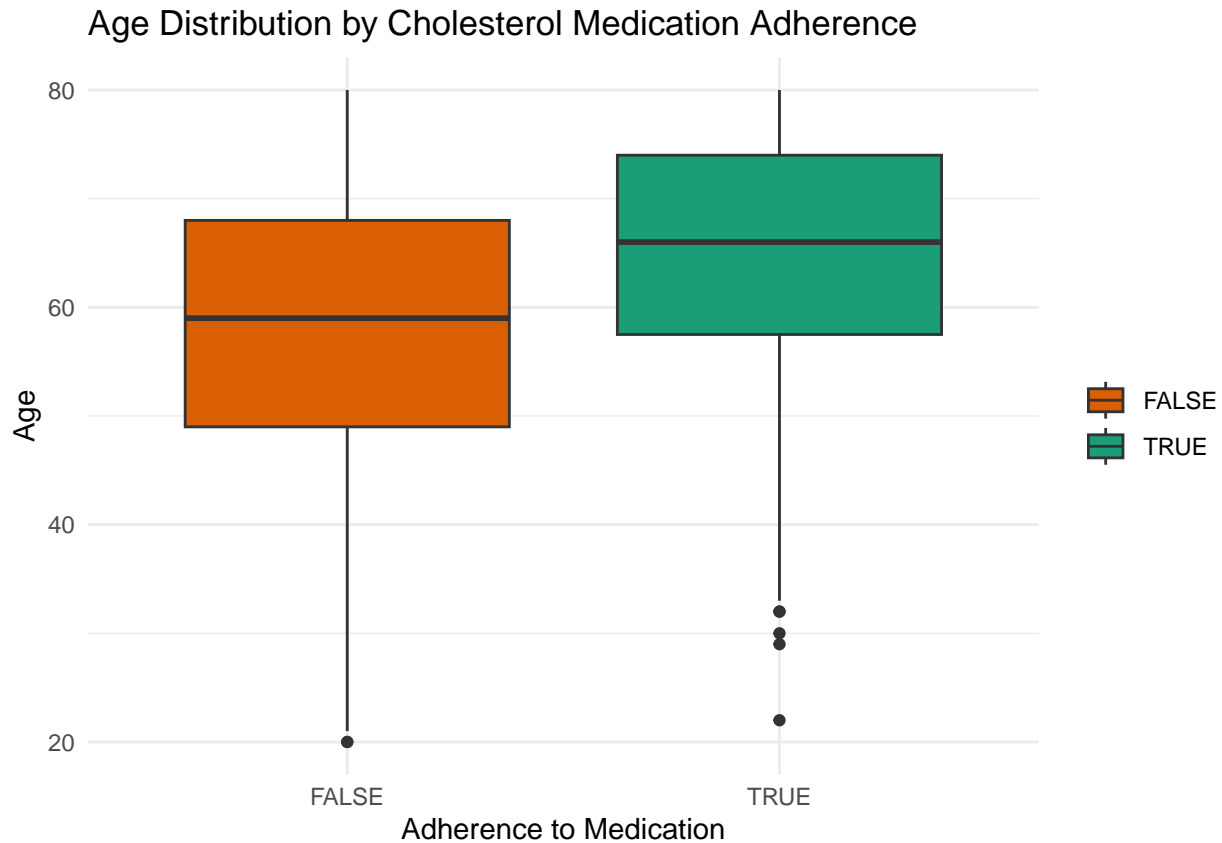
The resulting plots reveal several patterns related to adherence to cholesterol medication among adults:

1. **Adherence to Cholesterol Medication by Income Category**: The bar chart displays adherence to cholesterol medication across different income categories. The group with an income greater than 185% of the Federal Poverty Level (FPL) shows a larger proportion of individuals adhering to medication compared to the other groups, which aligns with the statistical significance found in the chi-squared test.

2. **Adherence to Cholesterol Medication by Health Insurance Status**: This chart demonstrates that individuals with government health insurance and drug coverage show higher adherence compared to those without drug coverage or with unknown coverage. The significant chi-squared test result is visually corroborated here, showing the importance of insurance and drug coverage on medication adherence.

3. **Adherence to Cholesterol Medication by Race/Ethnicity**: The distribution across race/ethnicity groups shows that Non-Hispanic White individuals have the highest count of adherence, followed by Non-Hispanic Black individuals. The chi-squared test result was not statistically significant, suggesting that the observed differences might be due to chance. However, the plot still shows notable differences in adherence rates that may warrant further investigation.

4. **Adherence to Cholesterol Medication by Gender**: The chart indicates that males have a slightly higher count of non-adherence than females. The chi-squared test for gender was significant, suggesting a potential difference in adherence behavior between genders.

5. **Adherence to Cholesterol Medication by Education Level**: This bar chart illustrates the relationship between education level and adherence to cholesterol medication. It shows that individuals with some college or an associate's degree tend to adhere more to their cholesterol medication regimen compared to other educational levels. However, this trend does not present a clear gradient with increasing education, as those with a high school diploma or GED show a lower adherence rate, which is

slightly less than those who didn't complete high school. Meanwhile, individuals with college degrees or higher show a moderate adherence rate. The chi-squared test did not find a statistically significant association between education level and adherence, suggesting that factors other than education level may play a more influential role in determining medication adherence.

6. **Adherence to Cholesterol Medication by Marital Status**: The visualization suggests that marital status may play a role in adherence to cholesterol medication. Individuals who are married or living with a partner show a higher adherence rate compared to those who have never been married or are widowed/divorced/separated. This is consistent with the chi-squared test indicating a significant association between marital status and medication adherence.

7. **Adherence to Cholesterol Medication by Alcohol Use**: The bar chart suggests that heavy alcohol users have a slightly lower adherence to cholesterol medication than those who do not engage in heavy drinking, as indicated by the larger proportion of non-adherence in the 'TRUE' category for heavy drinking. However, this association is not statistically significant at the 0.05 level, with a p-value just above the threshold.

8. **Adherence to Cholesterol Medication by Tobacco Use**: The chart illustrates adherence to cholesterol medication among different categories of tobacco users. Individuals who have never smoked show a higher proportion of adherence to medication, followed closely by past smokers, while current smokers have the lowest proportion of adherence. The chi-squared test result supports a significant association between tobacco use and medication adherence.

9. **Age Distribution by Cholesterol Medication Adherence**: The box plot shows a clear difference in the age distribution between those who adhere and those who do not, with the median age of adherent individuals being higher. This is consistent with the t-test results, which indicated a significant difference in age between the two groups.

Bivariate Analysis Visualization Interpretation

Our bivariate analysis visualizations comprehensively encapsulate factors like socioeconomic status, insurance coverage, demographic characteristics, and personal behaviors, including marital status, alcohol, and tobacco use. These visual representations are crucial for a quick and clear understanding of the data and for effectively sharing our findings.

Through these visualizations, we see that marital status continues to show a strong link with medication adherence, suggesting that being married or having a partner may contribute to better health practices. Tobacco use also appears to be an influential factor; individuals who have never smoked tend to adhere more to medication regimens, potentially indicative of overall healthier lifestyle choices. Alcohol use does not show a statistically significant association with adherence, but the pattern observed could be informative in a broader health context.

The bivariate analysis also sheds light on the role of education in medication adherence. The visualizations do not reveal a clear-cut trend across education levels, suggesting that higher education does not necessarily equate to higher adherence. This could imply that education level alone is not a straightforward predictor of adherence, and that it must be considered alongside other factors that may exert a more complex influence.

In summary, these visualizations underscore the intricate web of factors that interplay in the realm of medication adherence. They stress the importance of considering a range of variables, including demographic, socioeconomic, behavioral, and educational, to gain a full understanding. This comprehensive approach is vital for crafting effective interventions and policies to improve adherence rates and tackle health disparities. The next step in our analysis, a multivariate approach, will be critical to unravel the combined impact of these diverse factors on adherence to cholesterol medication.

Multivariate Analysis

Next, we'll conduct a logistic regression analysis to assess the relationship between family income and medication adherence, controlling for confounders.

```
# logistic regression model with all variables
lr_model <- glm(adherence ~ income_cat + ins_classif + race_6cat + sex + age +
                    educ_level + marital_status + alc_heavy + smoking_hx,
                family = binomial(link = "logit"),
                data = subset_1168_data)

# Summary of the model
summary(lr_model)
```

```
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + marital_status + alc_heavy + smoking_hx,
##     family = binomial(link = "logit"), data = subset_1168_data)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.703451  88.719758  -0.019   0.9847
## income_cat.L       0.154916   0.145265   1.066   0.2862
## income_cat.Q      -0.167241   0.175152  -0.955   0.3397
## ins_classif.L     -6.774035 172.673848  -0.039   0.9687
## ins_classif.Q      2.096812 187.926712   0.011   0.9911
## ins_classif.C      9.550539 293.732256   0.033   0.9741
## ins_classif^4     -6.707651 182.930096  -0.037   0.9707
## ins_classif^5     -9.280138 238.828741  -0.039   0.9690
## ins_classif^6      5.877962 306.505889   0.019   0.9847
## ins_classif^7      1.215256 381.627765   0.003   0.9975
## ins_classif^8     11.140162 291.906819   0.038   0.9696
## race_6cat.L        0.809558   0.350659   2.309   0.0210 *
## race_6cat.Q        0.473227   0.314953   1.503   0.1330
## race_6cat.C        0.320442   0.271889   1.179   0.2386
## race_6cat^4       -0.077656   0.230224  -0.337   0.7359
## race_6cat^5       -0.154706   0.161902  -0.956   0.3393
## sex.L              0.049766   0.111842   0.445   0.6563
## age                0.033213   0.007616   4.361 1.3e-05 ***
## educ_level.L      -0.074002   0.266525  -0.278   0.7813
## educ_level.Q      -0.012524   0.234105  -0.053   0.9573
## educ_level.C       0.078636   0.203349   0.387   0.6990
## educ_level^4       0.064196   0.171860   0.374   0.7087
## marital_status.L  -0.095941   0.186575  -0.514   0.6071
## marital_status.Q   0.083286   0.158708   0.525   0.5997
## alc_heavyTRUE      0.142977   0.261192   0.547   0.5841
## smoking_hx.L      -0.358103   0.153923  -2.326   0.0200 *
## smoking_hx.Q      -0.246001   0.139914  -1.758   0.0787 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

16

```
##     Null deviance: 1194.9  on 956  degrees of freedom
## Residual deviance: 1102.9  on 930  degrees of freedom
##   (211 observations deleted due to missingness)
## AIC: 1156.9
##
## Number of Fisher Scoring iterations: 13
```

The logistic regression model provides an estimate of the association between various predictor variables and the likelihood of adherence to cholesterol medication. Here's an interpretation of the key findings from the model output:

1. **Income Category**: The coefficients for income_cat.L (linear term) and income_cat.Q (quadratic term) are not statistically significant ($p > 0.05$), suggesting no clear association between income category and medication adherence based on the available data.

2. **Insurance Classification**: None of the insurance classification terms (linear through to the eighth polynomial term) are statistically significant, indicating that within this model, insurance classification does not have a statistically significant association with medication adherence.

3. **Race/Ethnicity**: The linear term race_6cat.L is significant ($p = 0.021$), suggesting that there may be differences in medication adherence across different racial/ethnic categories. However, higher-order terms are not significant, indicating that the relationship is not straightforward and may not follow a clear pattern across all categories.

4. **Sex**: The term sex.L (representing sex of the individual) is not significant, implying that sex does not have a statistically significant impact on medication adherence in this model.

5. **Age**: Age is significant ($p < 0.0001$), indicating a positive association with adherence to cholesterol medication. This suggests that as age increases, the likelihood of adhering to medication also increases.

6. **Education Level**: Education level terms are not statistically significant, suggesting no clear pattern or association between different levels of education and medication adherence.

7. **Marital Status**: The terms for marital status are not statistically significant, indicating that marital status, as categorized in this model, does not have a significant impact on medication adherence.

8. **Alcohol Use**: The term for heavy alcohol use (alc_heavyTRUE) is not statistically significant, suggesting that heavy alcohol use does not have a significant impact on medication adherence in this model.

9. **Tobacco Use**: The linear term for tobacco use (smoking_hx.L) is significant ($p = 0.020$), indicating that there is a significant association between tobacco use and medication adherence. The negative coefficient suggests that individuals with a history of tobacco use are less likely to adhere to their medication. The quadratic term (smoking_hx.Q) approaches significance ($p = 0.0787$), indicating a possible non-linear relationship between tobacco use history and medication adherence.

Overall, the model indicates that age and certain aspects of race/ethnicity and tobacco use history are associated with medication adherence, while other factors such as income, insurance classification, sex, education level, marital status, and heavy alcohol use are not significantly associated in this analysis. It's important to note that the lack of significance for some variables does not necessarily mean that they have no effect on medication adherence but rather that any effect is not detectable with this model and dataset. Additional analysis, perhaps with different modeling strategies or additional data, may be required to fully understand these relationships.

```r
# Calculate Odds Ratios and 95% Confidence Intervals
or <- exp(coef(lr_model))
# Wald Confidence Intervals and p-values
se <- sqrt(diag(vcov(lr_model)))
wald_ci_lower <- exp(coef(lr_model) - 1.96 * se)
wald_ci_upper <- exp(coef(lr_model) + 1.96 * se)
p_values <- summary(lr_model)$coefficients[, "Pr(>|z|)"]

# Create a data frame to nicely format the results
results <- data.frame(
  OR = exp(coef(lr_model)),
  LowerCI = wald_ci_lower,
  UpperCI = wald_ci_upper,
  PValue = p_values
)

# View the results
print(results)
```

```
##                          OR        LowerCI        UpperCI        PValue
## (Intercept)    1.820541e-01   5.500703e-77   6.025358e+74  9.846813e-01
## income_cat.L   1.167560e+00   8.782691e-01   1.552140e+00  2.862235e-01
## income_cat.Q   8.459955e-01   6.001719e-01   1.192506e+00  3.396614e-01
## ins_classif.L  1.143073e-03  1.188851e-150  1.099057e+144 9.687068e-01
## ins_classif.Q  8.140181e+00  8.794025e-160  7.534951e+160 9.910977e-01
## ins_classif.C  1.405226e+04  1.311602e-246  1.505533e+254 9.740618e-01
## ins_classif^4  1.221530e-03  2.364086e-159  6.311686e+152 9.707499e-01
## ins_classif^5  9.325828e-05  4.726709e-208  1.839992e+199 9.690045e-01
## ins_classif^6  3.570809e+02  4.463617e-259  2.856579e+263 9.846996e-01
## ins_classif^7  3.371158e+00   0.000000e+00            Inf 9.974592e-01
## ins_classif^8  6.888279e+04  2.301496e-244  2.061632e+253 9.695574e-01
## race_6cat.L    2.246914e+00   1.130056e+00   4.467588e+00  2.096148e-02
## race_6cat.Q    1.605166e+00   8.658184e-01   2.975865e+00  1.329592e-01
## race_6cat.C    1.377736e+00   8.085915e-01   2.347485e+00  2.385680e-01
## race_6cat^4    9.252826e-01   5.892556e-01   1.452931e+00  7.358860e-01
## race_6cat^5    8.566673e-01   6.237331e-01   1.176591e+00  3.392979e-01
## sex.L          1.051025e+00   8.441336e-01   1.308625e+00  6.563422e-01
## age            1.033771e+00   1.018454e+00   1.049318e+00  1.295125e-05
## educ_level.L   9.286697e-01   5.507960e-01   1.565784e+00  7.812768e-01
## educ_level.Q   9.875538e-01   6.241470e-01   1.562553e+00  9.573345e-01
## educ_level.C   1.081811e+00   7.262015e-01   1.611557e+00  6.989733e-01
## educ_level^4   1.066301e+00   7.613595e-01   1.493380e+00  7.087488e-01
## marital_status.L 9.085175e-01 6.302562e-01   1.309633e+00  6.070971e-01
## marital_status.Q 1.086853e+00 7.962983e-01   1.483425e+00  5.997386e-01
## alc_heavyTRUE  1.153704e+00   6.914538e-01   1.924976e+00  5.841019e-01
## smoking_hx.L   6.990014e-01   5.169587e-01   9.451489e-01  1.999191e-02
## smoking_hx.Q   7.819214e-01   5.943823e-01   1.028633e+00  7.870836e-02
```

```r
####### Uncomment this chunk to generate the multivariate results table #######
# # Load the required packages
# library(knitr)
# library(kableExtra)
#
```

```
# # Create a nice looking table to present the above ORs, CIs, and p-values
# nice_table <- kable(results,
#                     format = "html",  # Use "latex" for PDF output or "pipe" for Markdown or "html" f
#                     digits = 3,        # Number of decimal places
#                     align = 'c',       # Center align the columns
#                     caption = "Multivariate Analysis of Factors Associated with Adherence to Choleste
#   kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
#                 full_width = F,
#                 position = "center") %>%
#   column_spec(1, bold = T) %>%        # Make the OR column bold
#   column_spec(2:4, color = "blue") %>% # Color the CI columns blue
#   scroll_box(width = "100%", height = "2000px") # Add a scroll box if the table is too large
#
# # Print the table
# nice_table
#
# # To display this table outside of an R Markdown document, save it to an HTML file and open it in a w
# save_kable(nice_table, file = "Multivariate_Results_Table_1168_subset.html")
```

The table presents the odds ratios (OR), confidence intervals (LowerCI and UpperCI), and p-values (PValue) for the variables included in the logistic regression model predicting medication adherence. Here's the interpretation of some key findings:

1. **Income Category**: The linear term (income_cat.L) has an OR of 1.17, suggesting that as income increases, there is a modest increase in the odds of medication adherence, although this is not statistically significant (p = 0.286). The quadratic term (income_cat.Q) OR of 0.85 indicates a non-linear relationship between income and adherence, but again, it is not statistically significant (p = 0.340).

2. **Insurance Classification**: The coefficients for insurance classification are represented by polynomial terms, with very high ORs for some terms (e.g., ins_classif.C OR = 14,052.26) but also very high p-values, indicating non-significant results. These might be due to overfitting or multicollinearity issues when high-order polynomial terms are included.

3. **Race/Ethnicity**: The linear term (race_6cat.L) shows a significant relationship with adherence (OR = 2.25, p = 0.021), suggesting that racial/ethnic differences do influence medication adherence.

4. **Sex**: The OR for the linear term of sex (sex.L) is 1.05, which indicates a slight increase in the odds of adherence for one sex compared to the other. However, this effect is not statistically significant (p = 0.656), suggesting that in this model, sex does not have a significant impact on medication adherence when controlling for other factors. The way sex is coded in the model, likely as a binary variable, does not indicate which sex has higher odds of adherence; it only indicates that there's no strong evidence of a difference in adherence between the sexes in this dataset.

5. **Age**: A significant relationship is seen with age, where each additional year increases the odds of adherence by about 3.4% (OR = 1.03, p < 0.00001), indicating a stronger adherence among older individuals.

6. **Education Level**: The education level variables are not significant, with ORs close to 1 and p-values well above the traditional 0.05 threshold for significance.

7. **Marital Status**: No significant effects were observed for marital status, with ORs close to 1 and non-significant p-values.

8. **Alcohol Use**: Heavy alcohol use was not significantly associated with medication adherence (OR = 1.15, p = 0.584).

9. **Tobacco Use**: The linear term for smoking history (smoking_hx.L) shows a significant negative relationship with adherence (OR = 0.70, p = 0.020), suggesting that those with a history of smoking have lower odds of adhering to medication.

It's important to note that these results should be interpreted with caution, especially for the insurance classification where the extremely large ORs and wide confidence intervals indicate potential instability in the model for these variables. This could be due to the small number of cases in certain categories or other issues with the data. Further investigation and possibly model refinement are warranted to clarify these relationships.

Multivariate Analysis Interpretation

The multivariate analysis reveals a complex interplay of factors influencing adherence to cholesterol medication. Notably, age stands out as a significant predictor, with a slight increase in the likelihood of adherence as age increases (OR = 1.034, p < 0.0001). Tobacco use history also emerges as an important factor; those with no history of smoking are more likely to adhere to their medication regime (OR for smoking history linear term = 0.699, p = 0.02). Other factors, including income and insurance classification, though represented with polynomial terms, do not show a statistically significant effect on adherence in this model. Variables like sex, marital status, and alcohol use did not have a significant impact on medication adherence when controlling for other factors. This analysis underscores the importance of age and tobacco use in medication adherence while also suggesting that the relationships between adherence and socioeconomic factors may be nuanced and require further exploration.

Multivariate Analysis Visualization

```r
# First, we need to extract the model coefficients and their confidence intervals
model_coef <- coef(summary(lr_model))
model_or <- exp(cbind(OR = model_coef[, "Estimate"],
                      LowerCI = model_coef[, "Estimate"] - 1.96 * model_coef[, "Std. Error"],
                      UpperCI = model_coef[, "Estimate"] + 1.96 * model_coef[, "Std. Error"]))

# Convert to a data frame for plotting
model_or_df <- as.data.frame(model_or)
model_or_df$Variable <- rownames(model_or_df)
model_or_df$OR <- exp(model_or_df$OR)  # Convert log-odds to odds

# Reorder the variables based on the OR
model_or_df <- model_or_df[order(model_or_df$OR), ]

# Plotting using ggplot2
library(ggplot2)
lr_plot <- ggplot(model_or_df, aes(x = reorder(Variable, OR), y = OR)) +
  geom_point() +
  geom_errorbar(aes(ymin = exp(LowerCI), ymax = exp(UpperCI)), width = 0.2) +
  coord_flip() +  # Flip coordinates for horizontal layout
  xlab("Variable") + ylab("Odds Ratio (OR)") +
  theme_minimal() +
  geom_hline(yintercept = 1, linetype = "dashed", color = "blue")  # Reference line for OR = 1

# Print the plot
print(lr_plot)
```
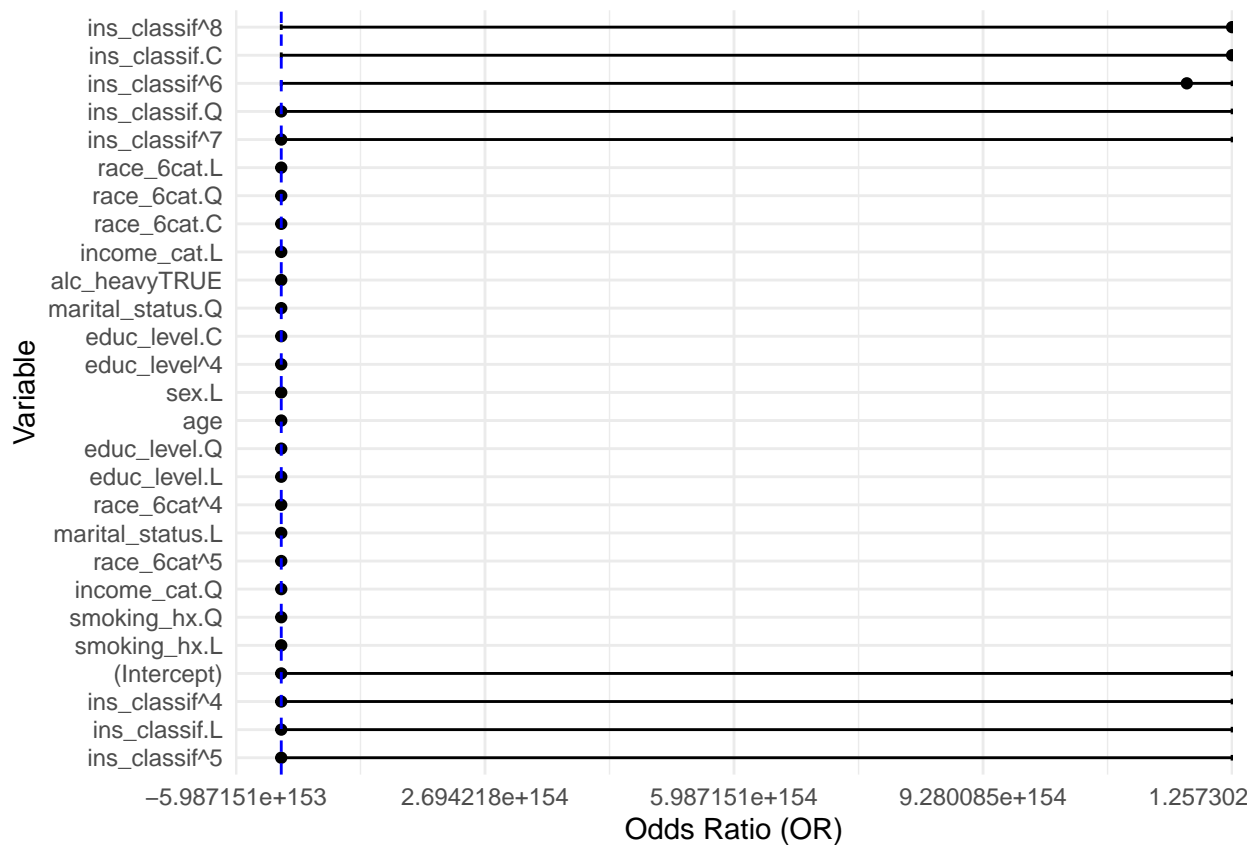
This plot is not informative as some ORs are really large and some CIs are really wide.

Stratified Analyses

```r
# Stratified analysis by income level
stratified_by_income <- subset_1168_data %>%
  split(.$income_cat) %>%
  lapply(function(data) {
    glm(adherence ~ ins_classif + race_6cat + sex + age + educ_level + marital_status + alc_heavy + smo
        family = binomial(link = "logit"),
        data = data)
  })

# Stratified analysis by race/ethnicity
stratified_by_race <- subset_1168_data %>%
  split(.$race_6cat) %>%
  lapply(function(data) {
    glm(adherence ~ income_cat + ins_classif + sex + age + educ_level + marital_status + alc_heavy + smo
        family = binomial(link = "logit"),
        data = data)
  })

# Stratified analysis by gender
stratified_by_gender <- subset_1168_data %>%
  split(.$sex) %>%
```

```r
  lapply(function(data) {
    glm(adherence ~ income_cat + ins_classif + race_6cat + age + educ_level + marital_status + alc_heavy
        family = binomial(link = "logit"),
        data = data)
  })

# Stratified analysis by insurance status with checks
stratified_by_insurance <- subset_1168_data %>%
  split(.$ins_classif) %>%
  lapply(function(data) {
    predictors <- c("income_cat", "race_6cat", "sex", "age", "educ_level", "marital_status", "alc_heavy"
    valid_predictors <- predictors[sapply(data[, predictors], function(x) length(unique(x)) > 1)]

    if(length(valid_predictors) == length(predictors)) {
      glm(adherence ~ income_cat + race_6cat + sex + age + educ_level + marital_status + alc_heavy + smo
          family = binomial(link = "logit"),
          data = data)
    } else {
      cat("Skipped model for", names(data), "due to insufficient data.\n")
      NULL
    }
  })
```

```
## Skipped model for id drop_flag collection_cycle collection_lang cholesterol_rx_ever adherence cancer
## Skipped model for id drop_flag collection_cycle collection_lang cholesterol_rx_ever adherence cancer
## Skipped model for id drop_flag collection_cycle collection_lang cholesterol_rx_ever adherence cancer
```

```r
stratified_by_insurance <- Filter(Negate(is.null), stratified_by_insurance)

# Stratified analysis by education level
stratified_by_education <- subset_1168_data %>%
  split(.$educ_level) %>%
  lapply(function(data) {
    glm(adherence ~ income_cat + ins_classif + race_6cat + sex + age + marital_status + alc_heavy + smol
        family = binomial(link = "logit"),
        data = data)
  })

# Stratified analysis by marital status
stratified_by_marital <- subset_1168_data %>%
  split(.$marital_status) %>%
  lapply(function(data) {
    glm(adherence ~ income_cat + ins_classif + race_6cat + sex + age + educ_level + alc_heavy + smoking_
        family = binomial(link = "logit"),
        data = data)
  })

# Stratified analysis by alcohol use
stratified_by_alcohol <- subset_1168_data %>%
  split(.$alc_heavy) %>%
  lapply(function(data) {
    glm(adherence ~ income_cat + ins_classif + race_6cat + sex + age + educ_level + marital_status + smc
        family = binomial(link = "logit"),
```

```
        data = data)
  })

# Stratified analysis by tobacco use
stratified_by_tobacco <- subset_1168_data %>%
  split(.$smoking_hx) %>%
  lapply(function(data) {
    glm(adherence ~ income_cat + ins_classif + race_6cat + sex + age + educ_level + marital_status + al
        family = binomial(link = "logit"),
        data = data)
  })

# Output the summaries of the stratified analyses
lapply(stratified_by_income, summary)
```

```
## $`Less than or equal to 135% of FPL`
##
## Call:
## glm(formula = adherence ~ ins_classif + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.82101  269.70924  -0.003   0.9976
## ins_classif.L    -15.65382  806.37825  -0.019   0.9845
## ins_classif.Q    -11.02111  650.47436  -0.017   0.9865
## ins_classif.C      1.21909  727.98897   0.002   0.9987
## ins_classif^4    -11.40486  670.66395  -0.017   0.9864
## ins_classif^5      3.84775  359.36994   0.011   0.9915
## race_6cat.L        0.55180    0.80926   0.682   0.4953
## race_6cat.Q        1.55111    0.75392   2.057   0.0396 *
## race_6cat.C        0.70785    0.64962   1.090   0.2759
## race_6cat^4        0.39658    0.52817   0.751   0.4527
## race_6cat^5        0.05017    0.34106   0.147   0.8831
## sex.L             -0.11916    0.22912  -0.520   0.6030
## age                0.02422    0.01358   1.784   0.0744 .
## educ_level.L       0.08415    0.47605   0.177   0.8597
## educ_level.Q      -0.29628    0.45385  -0.653   0.5139
## educ_level.C       0.23755    0.34570   0.687   0.4920
## educ_level^4       0.49694    0.30258   1.642   0.1005
## marital_status.L   0.09692    0.31469   0.308   0.7581
## marital_status.Q  -0.45844    0.28900  -1.586   0.1127
## alc_heavyTRUE     -0.34633    0.45979  -0.753   0.4513
## smoking_hx.L      -0.38090    0.26556  -1.434   0.1515
## smoking_hx.Q      -0.61538    0.28033  -2.195   0.0281 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 314.52  on 238  degrees of freedom
## Residual deviance: 273.15  on 217  degrees of freedom
```

```
##    (73 observations deleted due to missingness)
## AIC: 317.15
##
## Number of Fisher Scoring iterations: 15
##
##
## $`Between 135 and 185% of FPL`
##
## Call:
## glm(formula = adherence ~ ins_classif + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        3.01030  564.22706   0.005  0.99574
## ins_classif.L    -10.15260  908.61990  -0.011  0.99108
## ins_classif.Q      6.92844  908.43632   0.008  0.99391
## ins_classif.C     15.07790 1387.79904   0.011  0.99133
## ins_classif^4     -9.23527 1129.66301  -0.008  0.99348
## ins_classif^5     -3.62496 1739.65015  -0.002  0.99834
## ins_classif^6    -11.48963 1423.27754  -0.008  0.99356
## race_6cat.L       11.42865 1029.99518   0.011  0.99115
## race_6cat.Q        8.39136  940.25261   0.009  0.99288
## race_6cat.C        6.08449  642.31513   0.009  0.99244
## race_6cat^4        2.78522  325.71390   0.009  0.99318
## race_6cat^5       -0.45760  108.57248  -0.004  0.99664
## sex.L              0.15141    0.36008   0.420  0.67412
## age               -0.01036    0.02988  -0.347  0.72888
## educ_level.L      -0.42136    0.74608  -0.565  0.57224
## educ_level.Q      -0.21186    0.66022  -0.321  0.74829
## educ_level.C       1.22801    0.71555   1.716  0.08613 .
## educ_level^4      -0.11153    0.59590  -0.187  0.85154
## marital_status.L  -2.14439    0.70373  -3.047  0.00231 **
## marital_status.Q  -0.33013    0.52948  -0.624  0.53295
## alc_heavyTRUE      1.21182    1.02168   1.186  0.23558
## smoking_hx.L      -1.08794    0.52757  -2.062  0.03919 *
## smoking_hx.Q       0.04722    0.45821   0.103  0.91792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 182.98  on 157  degrees of freedom
## Residual deviance: 125.84  on 135  degrees of freedom
##   (44 observations deleted due to missingness)
## AIC: 171.84
##
## Number of Fisher Scoring iterations: 16
##
##
## $`Greater than 185% FPL`
##
## Call:
```

```
## glm(formula = adherence ~ ins_classif + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.933085 118.149931  -0.016    0.987
## ins_classif.L    -5.824724 190.831961  -0.031    0.976
## ins_classif.Q     1.728531 298.923405   0.006    0.995
## ins_classif.C     8.781189 351.778458   0.025    0.980
## ins_classif^4    -6.862218 229.714069  -0.030    0.976
## ins_classif^5   -10.321908 383.253196  -0.027    0.979
## ins_classif^6     5.343305 293.767891   0.018    0.985
## ins_classif^7     0.588657 508.220371   0.001    0.999
## ins_classif^8    11.351498 459.655104   0.025    0.980
## race_6cat.L       0.596829   0.445563   1.339    0.180
## race_6cat.Q       0.189879   0.395571   0.480    0.631
## race_6cat.C       0.115347   0.356457   0.324    0.746
## race_6cat^4      -0.313050   0.308497  -1.015    0.310
## race_6cat^5      -0.050777   0.216861  -0.234    0.815
## sex.L             0.050157   0.149986   0.334    0.738
## age               0.043847   0.010838   4.046 5.22e-05 ***
## educ_level.L     -0.532822   0.576723  -0.924    0.356
## educ_level.Q      0.269222   0.504381   0.534    0.594
## educ_level.C     -0.000425   0.413801  -0.001    0.999
## educ_level^4     -0.317710   0.299068  -1.062    0.288
## marital_status.L  0.241812   0.296815   0.815    0.415
## marital_status.Q  0.389099   0.241661   1.610    0.107
## alc_heavyTRUE     0.253105   0.373540   0.678    0.498
## smoking_hx.L     -0.296317   0.236122  -1.255    0.210
## smoking_hx.Q     -0.181407   0.200494  -0.905    0.366
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 692.43  on 559  degrees of freedom
## Residual deviance: 637.76  on 535  degrees of freedom
##   (94 observations deleted due to missingness)
## AIC: 687.76
##
## Number of Fisher Scoring iterations: 13
```

The stratified analysis by income level reveals distinct patterns of adherence to cholesterol medication across different income groups. Here's an interpretation of the results for each income category:

1. **Less than or equal to 135% of FPL (Federal Poverty Level)**:

   - **Race/Ethnicity Quadratic Component (race_6cat.Q)** shows a significant positive association with adherence, suggesting a complex, non-linear relationship between race/ethnicity and adherence in this income group.
   - **Age** shows a trend towards significance, indicating that older individuals in this income group might be more likely to adhere to medication, though this is not statistically significant at the conventional 0.05 level.

- **Tobacco Use Linear Component (smoking_hx.L)** is significantly negatively associated with adherence, implying that individuals who have never smoked are more likely to adhere to cholesterol medication.

2. **Between 135 and 185% of FPL**:
    - **Marital Status Linear Component (marital_status.L)** shows a significant negative association with adherence, suggesting that individuals who are not married (single, divorced, widowed) are less likely to adhere to medication.
    - **Tobacco Use Linear Component (smoking_hx.L)** is again negatively associated with adherence, reinforcing the trend seen in the lower income group.

3. **Greater than 185% FPL**:
    - **Age** is significantly positively associated with adherence, indicating that older individuals in this higher income group are more likely to adhere to cholesterol medication.
    - Other variables, including marital status and tobacco use, do not show significant associations in this income group.

These findings underscore the influence of socioeconomic status on health behavior and the complex interplay of various factors like race/ethnicity, marital status, and tobacco use in influencing medication adherence. The significant variables and their directions vary across income levels, highlighting the need for targeted approaches in promoting medication adherence.

```
lapply(stratified_by_race, summary)
```

```
## $`Mexican American`
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        1.01781  565.58329    0.002   0.9986
## income_cat.L      -0.78993    0.76450   -1.033   0.3015
## income_cat.Q       0.74746    0.69570    1.074   0.2826
## ins_classif.L      2.96922 1672.32035    0.002   0.9986
## ins_classif.Q      7.74476 1334.98387    0.006   0.9954
## ins_classif.C     14.05252 1538.53938    0.009   0.9927
## ins_classif^4     -6.95465 1434.00304   -0.005   0.9961
## ins_classif^5      6.18378  770.75432    0.008   0.9936
## sex.L             -1.09302    0.61951   -1.764   0.0777 .
## age                0.06218    0.03877    1.604   0.1088
## educ_level.L      -1.30862    1.10598   -1.183   0.2367
## educ_level.Q      -1.01264    1.07508   -0.942   0.3462
## educ_level.C       0.96702    1.42855    0.677   0.4985
## educ_level^4       0.49787    1.19759    0.416   0.6776
## marital_status.L  -2.41365    1.58718   -1.521   0.1283
## marital_status.Q  -1.16415    1.18764   -0.980   0.3270
## alc_heavyTRUE      0.38567    1.05384    0.366   0.7144
## smoking_hx.L      -1.31987    1.35050   -0.977   0.3284
## smoking_hx.Q      -0.09135    1.13336   -0.081   0.9358
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 74.095  on 55  degrees of freedom
## Residual deviance: 54.990  on 37  degrees of freedom
##   (16 observations deleted due to missingness)
## AIC: 92.99
##
## Number of Fisher Scoring iterations: 15
##
##
## $'Other Hispanic'
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.94804  565.58080  -0.003   0.9973
## income_cat.L       0.25789    0.55067   0.468   0.6396
## income_cat.Q       0.23265    0.52425   0.444   0.6572
## ins_classif.L    -15.79261 1672.31984  -0.009   0.9925
## ins_classif.Q    -11.13018 1334.98367  -0.008   0.9933
## ins_classif.C      1.60742 1538.53906   0.001   0.9992
## ins_classif^4    -11.72460 1434.00255  -0.008   0.9935
## ins_classif^5      3.52261  770.75359   0.005   0.9964
## sex.L             -0.22453    0.40965  -0.548   0.5836
## age                0.02911    0.02541   1.146   0.2519
## educ_level.L       0.06049    0.66528   0.091   0.9276
## educ_level.Q      -0.03413    0.65203  -0.052   0.9583
## educ_level.C       0.05466    0.54765   0.100   0.9205
## educ_level^4       0.27040    0.61327   0.441   0.6593
## marital_status.L  -1.88531    0.87082  -2.165   0.0304 *
## marital_status.Q  -0.95326    0.63108  -1.511   0.1309
## alc_heavyTRUE     -0.33218    0.72866  -0.456   0.6485
## smoking_hx.L       0.63800    0.71802   0.889   0.3742
## smoking_hx.Q       0.91020    0.55194   1.649   0.0991 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 123.81  on 97  degrees of freedom
## Residual deviance: 107.13  on 79  degrees of freedom
##   (30 observations deleted due to missingness)
## AIC: 145.13
##
## Number of Fisher Scoring iterations: 15
##
##
## $'Non-Hispanic White'
```

27

```
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.25637  221.79640  -0.006   0.9955
## income_cat.L       0.04194    0.23677   0.177   0.8594
## income_cat.Q      -0.56695    0.28046  -2.022   0.0432 *
## ins_classif.L     -7.78832  410.68499  -0.019   0.9849
## ins_classif.Q      0.50866  410.68494   0.001   0.9990
## ins_classif.C     13.68458  764.43041   0.018   0.9857
## ins_classif^4      0.66866  214.47362   0.003   0.9975
## ins_classif^5    -11.26109  645.45286  -0.017   0.9861
## ins_classif^6     -1.73247  982.83899  -0.002   0.9986
## ins_classif^7    -11.40514  636.04455  -0.018   0.9857
## sex.L              0.03082    0.17166   0.180   0.8575
## age                0.02954    0.01182   2.499   0.0125 *
## educ_level.L       0.17941    0.54272   0.331   0.7410
## educ_level.Q       0.20530    0.45819   0.448   0.6541
## educ_level.C       0.12737    0.36154   0.352   0.7246
## educ_level^4       0.04075    0.26744   0.152   0.8789
## marital_status.L   0.38231    0.31793   1.202   0.2292
## marital_status.Q   0.41111    0.24945   1.648   0.0993 .
## alc_heavyTRUE      0.32828    0.40112   0.818   0.4131
## smoking_hx.L      -0.46252    0.24339  -1.900   0.0574 .
## smoking_hx.Q      -0.48389    0.20484  -2.362   0.0182 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 530.61  on 439  degrees of freedom
## Residual deviance: 482.21  on 419  degrees of freedom
##   (68 observations deleted due to missingness)
## AIC: 524.21
##
## Number of Fisher Scoring iterations: 14
##
##
## $'Non-Hispanic Black'
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -3.80163  126.11281  -0.030   0.9760
## income_cat.L       0.63731    0.28240   2.257   0.0240 *
## income_cat.Q       0.58192    0.42408   1.372   0.1700
```

```
## ins_classif.L      -3.35857  166.82562  -0.020   0.9839
## ins_classif.Q       3.89426  288.94722   0.013   0.9892
## ins_classif.C       6.46654  360.37942   0.018   0.9857
## ins_classif^4      -0.45606   71.14046  -0.006   0.9949
## ins_classif^5      -8.28139  481.57647  -0.017   0.9863
## ins_classif^6      -8.41571  435.60275  -0.019   0.9846
## sex.L               0.47168    0.23123   2.040   0.0414 *
## age                 0.02775    0.01813   1.530   0.1259
## educ_level.L        0.08391    0.99564   0.084   0.9328
## educ_level.Q       -1.41304    0.86407  -1.635   0.1020
## educ_level.C        0.86248    0.58424   1.476   0.1399
## educ_level^4        0.18659    0.39284   0.475   0.6348
## marital_status.L   -0.49967    0.33336  -1.499   0.1339
## marital_status.Q   -0.15794    0.30654  -0.515   0.6064
## alc_heavyTRUE       0.51273    0.66579   0.770   0.4412
## smoking_hx.L       -0.51184    0.29833  -1.716   0.0862 .
## smoking_hx.Q       -0.30766    0.29605  -1.039   0.2987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 295.46  on 221   degrees of freedom
## Residual deviance: 255.30  on 202   degrees of freedom
##   (66 observations deleted due to missingness)
## AIC: 295.3
##
## Number of Fisher Scoring iterations: 13
##
##
## $'Non-Hispanic Asian'
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients: (1 not defined because of singularities)
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      11.70143 2549.02769   0.005  0.99634
## income_cat.L      2.06110    0.90111   2.287  0.02218 *
## income_cat.Q     -1.32521    1.16234  -1.140  0.25424
## ins_classif.L    -3.67217  902.45190  -0.004  0.99675
## ins_classif.Q    -4.65516 3295.28369  -0.001  0.99887
## ins_classif.C     3.01898 2251.10749   0.001  0.99893
## ins_classif^4     6.76800 2853.79919   0.002  0.99811
## ins_classif^5   -11.81942 4756.33162  -0.002  0.99802
## sex.L            -0.72436    0.67110  -1.079  0.28043
## age               0.13427    0.04155   3.232  0.00123 **
## educ_level.L    -11.12030 2114.78774  -0.005  0.99580
## educ_level.Q     11.11002 1787.32184   0.006  0.99504
## educ_level.C     -8.32302 1057.39434  -0.008  0.99372
## educ_level^4      1.25611  399.65809   0.003  0.99749
## marital_status.L 13.79490 3803.17656   0.004  0.99711
```

```
## marital_status.Q   -6.22225 3512.28456  -0.002  0.99859
## alc_heavyTRUE            NA          NA     NA       NA
## smoking_hx.L      -0.89808     0.87115  -1.031  0.30258
## smoking_hx.Q       0.07083     0.80019   0.089  0.92947
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 126.719  on 103  degrees of freedom
## Residual deviance:  73.258  on  86  degrees of freedom
##   (20 observations deleted due to missingness)
## AIC: 109.26
##
## Number of Fisher Scoring iterations: 18
##
##
## $'Other Race (including multiracial)'
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.514e+00  1.265e+05   0.000    1.000
## income_cat.L     1.615e+01  4.412e+05   0.000    1.000
## income_cat.Q    -2.407e+01  2.556e+05   0.000    1.000
## ins_classif.L   -3.506e+01  1.992e+05   0.000    1.000
## ins_classif.Q   -2.358e+01  1.692e+05   0.000    1.000
## ins_classif.C    1.298e+01  3.983e+05   0.000    1.000
## ins_classif^4    2.112e+01  3.013e+05   0.000    1.000
## sex.L            1.295e+00  1.247e+00   1.038    0.299
## age              5.692e-02  2.103e-01   0.271    0.787
## educ_level.L     1.789e+01  2.851e+04   0.001    0.999
## educ_level.Q     4.576e+00  2.267e+04   0.000    1.000
## educ_level.C    -3.958e+00  3.382e+04   0.000    1.000
## educ_level^4    -4.158e+01  2.532e+04  -0.002    0.999
## marital_status.L -3.741e+01  3.830e+04  -0.001    0.999
## marital_status.Q -1.985e+01  2.211e+04  -0.001    0.999
## alc_heavyTRUE   -9.426e-01  1.421e+04   0.000    1.000
## smoking_hx.L     1.904e+01  4.409e+05   0.000    1.000
## smoking_hx.Q     2.697e+01  2.547e+05   0.000    1.000
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 32.800  on 36   degrees of freedom
## Residual deviance: 11.915  on 19  degrees of freedom
##   (11 observations deleted due to missingness)
## AIC: 47.915
##
## Number of Fisher Scoring iterations: 21
```

The results from the stratified analysis by race/ethnicity for adherence to cholesterol medication reveal varied patterns across different groups. Here's an interpretation of the results for each subgroup:

Mexican American

- **Significant Factors**: None of the variables show statistical significance, suggesting that within this subgroup, the included factors might not strongly predict medication adherence.
- **Trends**: There is a negative association for the linear term of sex (sex.L), indicating a potential gender-related trend in adherence, but this is not statistically significant.
- **Model Fit**: The model does not indicate a strong fit given the lack of significant predictors and a high AIC value.

Other Hispanic

- **Significant Factors**: Marital status (marital_status.L) shows a significant negative association, suggesting that marital status might play a role in medication adherence within this subgroup.
- **Trends**: There's a marginal trend in the quadratic term of smoking history (smoking_hx.Q), hinting at a complex relationship between smoking habits and adherence, but this isn't statistically significant.
- **Model Fit**: Similar to the Mexican American group, the model does not show a strong fit due to a lack of significant predictors.

Non-Hispanic White

- **Significant Factors**: Income (income_cat.Q) and age show significant negative and positive associations, respectively, indicating these are important predictors of adherence in this subgroup.
- **Trends**: There are marginal trends in smoking history, both linear and quadratic terms, suggesting a nuanced relationship between smoking and adherence.
- **Model Fit**: The presence of significant predictors implies a better model fit compared to the previous groups.

Non-Hispanic Black

- **Significant Factors**: Income (income_cat.L) and sex (sex.L) are significant predictors, indicating that these factors are influential in medication adherence within this subgroup.
- **Trends**: There's a marginal negative trend in the linear term of smoking history.
- **Model Fit**: The model shows a better fit with significant predictors compared to other groups.

Non-Hispanic Asian

- **Significant Factors**: Income (income_cat.L) and age are significant predictors of adherence, with age showing a positive association.
- **Trends**: The model couldn't estimate the association for alcohol use due to data limitations (indicated by NA).
- **Model Fit**: The significant predictors indicate a better model fit, although some coefficients could not be estimated.

Other Race (including multiracial)

- **Significant Factors**: No significant predictors are found in this group.

- **Trends**: Large standard errors and non-significant p-values suggest that the model may not be well specified for this group, or there may be insufficient data.
- **Model Fit**: The lack of significant predictors and the presence of large standard errors suggest a poor model fit.

These results underscore the importance of considering race/ethnicity in understanding medication adherence patterns, as different factors appear to be influential in different subgroups. The variability in significance and trends across these groups highlights the complexity of factors influencing medication adherence and the necessity of tailored approaches in different racial/ethnic contexts.

```
lapply(stratified_by_gender, summary)
```

```
## $Male
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     age + educ_level + marital_status + alc_heavy + smoking_hx,
##     family = binomial(link = "logit"), data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -3.65524   87.39181  -0.042 0.966637
## income_cat.L       0.16841    0.20661   0.815 0.415005
## income_cat.Q      -0.22004    0.26598  -0.827 0.408078
## ins_classif.L     -3.91367  115.60603  -0.034 0.972994
## ins_classif.Q      4.58563  200.23417   0.023 0.981729
## ins_classif.C      6.07306  249.73554   0.024 0.980599
## ins_classif^4     -1.20339   49.29668  -0.024 0.980525
## ins_classif^5     -8.70295  333.72289  -0.026 0.979195
## ins_classif^6     -7.73861  301.86374  -0.026 0.979548
## race_6cat.L        0.16570    0.47762   0.347 0.728635
## race_6cat.Q        0.86787    0.42663   2.034 0.041924 *
## race_6cat.C        0.18940    0.36457   0.520 0.603395
## race_6cat^4       -0.13191    0.30494  -0.433 0.665319
## race_6cat^5       -0.27954    0.22197  -1.259 0.207904
## age                0.03783    0.01037   3.648 0.000264 ***
## educ_level.L       0.21149    0.34417   0.614 0.538899
## educ_level.Q      -0.09091    0.30198  -0.301 0.763378
## educ_level.C       0.02632    0.28457   0.093 0.926299
## educ_level^4      -0.03908    0.24487  -0.160 0.873188
## marital_status.L   0.45841    0.28964   1.583 0.113489
## marital_status.Q   0.42915    0.25578   1.678 0.093380 .
## alc_heavyTRUE     -0.05739    0.33908  -0.169 0.865598
## smoking_hx.L      -0.27666    0.21372  -1.295 0.195487
## smoking_hx.Q      -0.19544    0.19042  -1.026 0.304723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 626.33  on 491  degrees of freedom
## Residual deviance: 570.49  on 468  degrees of freedom
##   (114 observations deleted due to missingness)
```

```
## AIC: 618.49
##
## Number of Fisher Scoring iterations: 13
##
##
## $Female
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     age + educ_level + marital_status + alc_heavy + smoking_hx,
##     family = binomial(link = "logit"), data = data)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.808214 120.014875  -0.015  0.98798
## income_cat.L       0.107996   0.216428   0.499  0.61779
## income_cat.Q      -0.049186   0.241526  -0.204  0.83863
## ins_classif.L     -6.254242 197.014646  -0.032  0.97468
## ins_classif.Q      2.040376 300.309153   0.007  0.99458
## ins_classif.C      9.228393 360.399602   0.026  0.97957
## ins_classif^4     -7.621594 234.398261  -0.033  0.97406
## ins_classif^5     -9.870156 384.854379  -0.026  0.97954
## ins_classif^6      5.166759 308.369097   0.017  0.98663
## ins_classif^7      0.949024 516.240354   0.002  0.99853
## ins_classif^8     11.129920 462.032186   0.024  0.98078
## race_6cat.L        1.504059   0.565302   2.661  0.00780 **
## race_6cat.Q        0.011831   0.506082   0.023  0.98135
## race_6cat.C        0.453378   0.452282   1.002  0.31614
## race_6cat^4       -0.018814   0.387469  -0.049  0.96127
## race_6cat^5       -0.001539   0.261450  -0.006  0.99530
## age                0.035658   0.012233   2.915  0.00356 **
## educ_level.L      -0.490528   0.457771  -1.072  0.28392
## educ_level.Q       0.232058   0.396936   0.585  0.55880
## educ_level.C      -0.061131   0.319764  -0.191  0.84839
## educ_level^4       0.177636   0.251749   0.706  0.48043
## marital_status.L  -0.697915   0.271785  -2.568  0.01023 *
## marital_status.Q  -0.193005   0.215090  -0.897  0.36955
## alc_heavyTRUE      0.468617   0.447045   1.048  0.29452
## smoking_hx.L      -0.347353   0.238161  -1.458  0.14471
## smoking_hx.Q      -0.287334   0.218310  -1.316  0.18812
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 567.25  on 464  degrees of freedom
## Residual deviance: 508.18  on 439  degrees of freedom
##   (97 observations deleted due to missingness)
## AIC: 560.18
##
## Number of Fisher Scoring iterations: 13
```

The results from the stratified analysis by gender for adherence to cholesterol medication present distinct patterns between males and females. Here's an interpretation of the results for each gender:

Male

- **Significant Factors**: Race (race_6cat.Q) and age are significant predictors, with age showing a strong positive association with adherence.
- **Trends**: The model suggests a marginal negative trend in sex (sex.L), indicating potential differences in adherence based on gender, but this is not statistically significant.
- **Model Fit**: The presence of significant predictors suggests a reasonable model fit, although several predictors are not significant, indicating a complex relationship between these factors and adherence.

Female

- **Significant Factors**: Race (race_6cat.L) and marital status (marital_status.L) show significant associations with adherence. The positive association with race suggests certain racial/ethnic groups may have higher adherence rates. Negative association with marital status indicates that certain marital statuses might be associated with lower adherence.
- **Trends**: Age is also a significant positive predictor, similar to males. The model does not show significant associations for other factors like income, insurance, and educational level.
- **Model Fit**: The model shows a reasonable fit, with significant predictors indicating a better understanding of adherence patterns among females.

In summary, the stratified analysis by gender reveals that certain factors like race and age play a significant role in predicting adherence to cholesterol medication, but this varies between males and females. The results highlight the need for gender-specific approaches in understanding and improving medication adherence. The lack of significant associations for some variables suggests the complexity of factors influencing adherence and the possibility of other unmeasured variables playing a role.

```
lapply(stratified_by_insurance, summary)
```

```
## $`Both private and government, with drug coverage`
##
## Call:
## glm(formula = adherence ~ income_cat + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        1.88267  211.89717   0.009    0.993
## income_cat.L       0.02003    0.37737   0.053    0.958
## income_cat.Q      -0.18437    0.40811  -0.452    0.651
## race_6cat.L        0.25527    0.94100   0.271    0.786
## race_6cat.Q        0.07837    0.77413   0.101    0.919
## race_6cat.C       -0.76175    0.88656  -0.859    0.390
## race_6cat^4       -0.68991    0.80537  -0.857    0.392
## race_6cat^5       -0.73080    0.50524  -1.446    0.148
## sex.L             -0.07720    0.25815  -0.299    0.765
## age                0.03215    0.02003   1.605    0.108
## educ_level.L      -9.66075  670.06250  -0.014    0.988
## educ_level.Q       8.08183  566.30616   0.014    0.989
## educ_level.C      -4.64517  335.03146  -0.014    0.989
## educ_level^4       1.76142  126.63050   0.014    0.989
## marital_status.L  -0.52774    0.61767  -0.854    0.393
## marital_status.Q   0.06784    0.40413   0.168    0.867
```

```
## alc_heavyTRUE         0.83577    0.90389   0.925     0.355
## smoking_hx.L         -0.17746    0.37986  -0.467     0.640
## smoking_hx.Q         -0.46779    0.34100  -1.372     0.170
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 251.23  on 230  degrees of freedom
## Residual deviance: 234.23  on 212  degrees of freedom
##   (26 observations deleted due to missingness)
## AIC: 272.23
##
## Number of Fisher Scoring iterations: 15
##
##
## $`Both private and government, without drug coverage`
##
## Call:
## glm(formula = adherence ~ income_cat + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.904e+03  3.493e+06  -0.001    0.999
## income_cat.L     -6.431e+01  2.996e+05   0.000    1.000
## income_cat.Q     -3.764e+02  3.875e+07   0.000    1.000
## race_6cat.L       6.405e+02  3.191e+07   0.000    1.000
## race_6cat.Q       2.019e+02  2.373e+07   0.000    1.000
## race_6cat.C      -2.052e+02  1.067e+07   0.000    1.000
## sex.L            -2.119e+02  3.649e+05  -0.001    1.000
## age               4.270e+01  3.711e+04   0.001    0.999
## educ_level.L      4.231e+02  7.503e+07   0.000    1.000
## educ_level.Q     -2.679e+02  1.269e+07   0.000    1.000
## educ_level.C      3.668e+02  5.289e+05   0.001    0.999
## educ_level^4      1.875e+02  2.270e+07   0.000    1.000
## marital_status.L -2.987e+02  7.454e+05   0.000    1.000
## marital_status.Q -5.799e+01  3.875e+07   0.000    1.000
## alc_heavyTRUE     2.735e+02  4.746e+07   0.000    1.000
## smoking_hx.L     -4.764e+02  3.356e+07   0.000    1.000
## smoking_hx.Q      1.598e+02  1.938e+07   0.000    1.000
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2.0450e+01  on 20  degrees of freedom
## Residual deviance: 2.4469e-09  on  4  degrees of freedom
##   (4 observations deleted due to missingness)
## AIC: 34
##
## Number of Fisher Scoring iterations: 25
##
##
## $`Government, with drug coverage`
##
## Call:
```

```
## glm(formula = adherence ~ income_cat + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.38669    0.75030  -1.848 0.064575 .
## income_cat.L        0.07862    0.21600   0.364 0.715858
## income_cat.Q       -0.60853    0.31718  -1.919 0.055044 .
## race_6cat.L         1.93961    0.74477   2.604 0.009206 **
## race_6cat.Q         0.58419    0.66418   0.880 0.379096
## race_6cat.C         0.72777    0.55256   1.317 0.187810
## race_6cat^4        -0.30772    0.43408  -0.709 0.478388
## race_6cat^5        -0.18084    0.28371  -0.637 0.523868
## sex.L               0.05099    0.18409   0.277 0.781806
## age                 0.04249    0.01137   3.738 0.000185 ***
## educ_level.L        0.35303    0.39196   0.901 0.367749
## educ_level.Q       -0.27449    0.34842  -0.788 0.430812
## educ_level.C        0.43377    0.30027   1.445 0.148571
## educ_level^4        0.08359    0.26187   0.319 0.749563
## marital_status.L    0.02619    0.27035   0.097 0.922839
## marital_status.Q    0.10057    0.23861   0.421 0.673411
## alc_heavyTRUE       0.02184    0.39522   0.055 0.955937
## smoking_hx.L       -0.19418    0.23885  -0.813 0.416223
## smoking_hx.Q       -0.31964    0.22875  -1.397 0.162320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 448.22  on 371   degrees of freedom
## Residual deviance: 400.47  on 353   degrees of freedom
##   (48 observations deleted due to missingness)
## AIC: 438.47
##
## Number of Fisher Scoring iterations: 5
##
##
## $'Government, without drug coverage'
##
## Call:
## glm(formula = adherence ~ income_cat + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        3.036e+02  3.831e+06   0.000    1.000
## income_cat.L      -5.671e+01  1.542e+07   0.000    1.000
## income_cat.Q       1.138e+02  8.915e+06   0.000    1.000
## race_6cat.L        5.218e+01  4.641e+07   0.000    1.000
## race_6cat.Q       -1.648e+02  2.577e+07   0.000    1.000
## race_6cat.C       -3.745e+01  1.277e+07   0.000    1.000
## race_6cat^4        1.257e+02  1.723e+07   0.000    1.000
```

```
## race_6cat^5       -3.247e+01  6.876e+06   0.000     1.000
## sex.L              4.012e+01  1.183e+07   0.000     1.000
## age               -5.534e+00  6.876e+03  -0.001     0.999
## educ_level.L       8.136e-01  1.058e+07   0.000     1.000
## educ_level.Q       1.092e+02  1.168e+07   0.000     1.000
## educ_level.C      -9.723e+01  5.293e+06   0.000     1.000
## educ_level^4       7.917e+01  1.843e+07   0.000     1.000
## marital_status.L -4.992e+01  1.183e+07   0.000     1.000
## marital_status.Q -1.021e+02  6.832e+06   0.000     1.000
## alc_heavyTRUE      4.295e+01  1.673e+07   0.000     1.000
## smoking_hx.L      -9.744e+01  4.376e+05   0.000     1.000
## smoking_hx.Q      -9.024e+01  2.687e+05   0.000     1.000
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.1054e+01  on 36  degrees of freedom
## Residual deviance: 2.1352e-09  on 18  degrees of freedom
##   (7 observations deleted due to missingness)
## AIC: 38
##
## Number of Fisher Scoring iterations: 25
##
##
## $'Private, with drug coverage'
##
## Call:
## glm(formula = adherence ~ income_cat + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.90579    0.89953  -1.007    0.3140
## income_cat.L     0.35168    0.33526   1.049    0.2942
## income_cat.Q     0.15740    0.36185   0.435    0.6636
## race_6cat.L      0.53372    0.53762   0.993    0.3208
## race_6cat.Q      0.50924    0.49808   1.022    0.3066
## race_6cat.C      0.50213    0.42219   1.189    0.2343
## race_6cat^4     -0.21548    0.36479  -0.591    0.5547
## race_6cat^5      0.02348    0.27330   0.086    0.9315
## sex.L            0.05975    0.20695   0.289    0.7728
## age              0.02623    0.01557   1.685    0.0920 .
## educ_level.L    -1.42490    0.74911  -1.902    0.0572 .
## educ_level.Q     0.85920    0.63964   1.343    0.1792
## educ_level.C    -0.68681    0.50466  -1.361    0.1735
## educ_level^4     0.30113    0.37295   0.807    0.4194
## marital_status.L  0.05780  0.34053   0.170    0.8652
## marital_status.Q -0.07460  0.35691  -0.209    0.8344
## alc_heavyTRUE    0.10644    0.44844   0.237    0.8124
## smoking_hx.L    -0.57821    0.30670  -1.885    0.0594 .
## smoking_hx.Q    -0.08828    0.26540  -0.333    0.7394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 372.87  on 272  degrees of freedom
## Residual deviance: 353.50  on 254  degrees of freedom
##   (30 observations deleted due to missingness)
## AIC: 391.5
##
## Number of Fisher Scoring iterations: 4
##
##
## $`Private, without drug coverage`
##
## Call:
## glm(formula = adherence ~ income_cat + race_6cat + sex + age +
##     educ_level + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients: (2 not defined because of singularities)
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -7394.6 38958261.2       0        1
## income_cat.L        2133.2 10905329.3       0        1
## income_cat.Q        1398.6  7334466.3       0        1
## race_6cat.L         -194.0  1044868.9       0        1
## race_6cat.Q         1366.6  7195011.7       0        1
## race_6cat.C         7179.2 37735525.3       0        1
## race_6cat^4        -2603.5 13752994.5       0        1
## sex.L               1265.5  6696001.8       0        1
## age                  153.4   808198.2       0        1
## educ_level.L        1086.2  5804335.9       0        1
## educ_level.Q       -2684.4 14060757.6       0        1
## educ_level.C        -971.8  5004959.5       0        1
## marital_status.L    3145.6 16533531.0       0        1
## marital_status.Q     313.1  1601899.3       0        1
## alc_heavyTRUE        255.7  1331513.8       0        1
## smoking_hx.L            NA         NA      NA       NA
## smoking_hx.Q            NA         NA      NA       NA
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.0190e+01  on 14  degrees of freedom
## Residual deviance: 2.3655e-10  on  0  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 30
##
## Number of Fisher Scoring iterations: 24
```

The results from the stratified analysis by insurance type for adherence to cholesterol medication show varied findings across different insurance categories. Here's an interpretation of the results for each category:

Both Private and Government, with Drug Coverage

- **Significant Factors**: None of the factors are statistically significant in predicting adherence.

- **Model Fit**: The model does not show a strong association between the included factors and medication adherence for this group. The high AIC value and large standard errors suggest the model may not be well-fitted to the data.

Both Private and Government, without Drug Coverage

- **Model Issues**: This model presents extremely large standard errors and estimates, which suggests issues with the data or model fit. The results are not meaningful for interpretation.
- **Data Limitations**: The results indicate possible limitations in the data, such as small sample size or lack of variation in predictor variables for this subgroup.

Government, with Drug Coverage

- **Significant Factors**: Race (race_6cat.L) and age are significant predictors. A positive coefficient for age indicates higher likelihood of adherence as age increases.
- **Model Fit**: The model seems better fitted for this group compared to the previous categories, as indicated by more reasonable standard errors and significant predictors.

Government, without Drug Coverage

- **Model Issues**: Like the model for "Both private and government, without drug coverage," this model also shows extremely large standard errors and estimates, suggesting issues with the data or model fit for this group.
- **Data Limitations**: This result suggests potential data limitations or insufficient sample size for meaningful analysis in this subgroup.

Private, with Drug Coverage

- **Trends**: No significant predictors were found. However, there's a marginal negative trend with education level (educ_level.L) and a positive trend with age.
- **Model Fit**: The model has better fit compared to some other categories but lacks strong predictive power.

Private, without Drug Coverage

- **Model Issues**: This model faces similar issues as seen in the "Government, without drug coverage" and "Both private and government, without drug coverage" categories, with extremely large estimates and standard errors, rendering the results unreliable.

In summary, the stratified analysis by insurance type reveals considerable variation in the model's predictive power across different insurance groups. While age and race emerge as significant predictors in some categories, the overall effectiveness of the model varies, with several categories showing potential data limitations or issues with model fit. This variation highlights the complexity of the relationship between insurance type and medication adherence, necessitating a nuanced approach for each category.

```
lapply(stratified_by_education, summary)
```

```
## $‘Less than 9th grade‘
##
## Call:
```

```
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -3.163e+00  7.245e+03   0.000    1.000
## income_cat.L        2.553e+01  4.913e+03   0.005    0.996
## income_cat.Q        1.471e+01  2.837e+03   0.005    0.996
## ins_classif.L      -4.754e+01  1.538e+04  -0.003    0.998
## ins_classif.Q       5.326e+01  9.226e+03   0.006    0.995
## ins_classif.C       2.309e+00  1.879e+04   0.000    1.000
## ins_classif^4       2.197e+01  1.966e+04   0.001    0.999
## ins_classif^5       9.154e+00  1.084e+04   0.001    0.999
## race_6cat.L         5.793e+01  1.486e+04   0.004    0.997
## race_6cat.Q         7.438e+01  1.511e+04   0.005    0.996
## race_6cat.C         4.311e+01  1.118e+04   0.004    0.997
## race_6cat^4        -1.475e+01  7.393e+03  -0.002    0.998
## race_6cat^5        -3.686e+01  7.402e+03  -0.005    0.996
## sex.L              -4.685e-01  9.686e-01  -0.484    0.629
## age                 1.192e-01  7.591e-02   1.570    0.116
## marital_status.L   -1.361e+01  3.334e+03  -0.004    0.997
## marital_status.Q   -8.078e+00  1.925e+03  -0.004    0.997
## alc_heavyTRUE       1.974e+01  4.715e+03   0.004    0.997
## smoking_hx.L       -5.427e+01  1.319e+04  -0.004    0.997
## smoking_hx.Q       -3.149e+01  7.613e+03  -0.004    0.997
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 58.630  on 48  degrees of freedom
## Residual deviance: 23.645  on 29  degrees of freedom
##   (16 observations deleted due to missingness)
## AIC: 63.645
##
## Number of Fisher Scoring iterations: 20
##
##
## $'9th-11th grade'
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          1.52965  928.24908   0.002   0.9987
## income_cat.L         0.95804    0.55355   1.731   0.0835 .
## income_cat.Q        -0.46188    0.76817  -0.601   0.5477
## ins_classif.L      -16.91233 1604.66574  -0.011   0.9916
## ins_classif.Q       -2.62084  952.85198  -0.003   0.9978
## ins_classif.C       17.52205 2002.50459   0.009   0.9930
## ins_classif^4       -1.60291 1852.05818  -0.001   0.9993
## ins_classif^5        9.67652 2213.04001   0.004   0.9965
```

```
## ins_classif^6          9.23157 2430.46480   0.004   0.9970
## ins_classif^7          0.10923 1475.72076   0.000   0.9999
## race_6cat.L            8.95921 2364.27039   0.004   0.9970
## race_6cat.Q            7.40926 2158.27358   0.003   0.9973
## race_6cat.C            6.37946 1474.38169   0.004   0.9965
## race_6cat^4            2.88134  747.64829   0.004   0.9969
## race_6cat^5            1.56680  249.21663   0.006   0.9950
## sex.L                 -0.18935    0.39576  -0.478   0.6323
## age                    0.04760    0.02262   2.105   0.0353 *
## marital_status.L      -0.33171    0.55325  -0.600   0.5488
## marital_status.Q      -0.15375    0.48226  -0.319   0.7499
## alc_heavyTRUE          0.03472    0.80232   0.043   0.9655
## smoking_hx.L           0.20722    0.46819   0.443   0.6581
## smoking_hx.Q          -0.22351    0.50355  -0.444   0.6571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 131.45  on 103  degrees of freedom
## Residual deviance: 101.28  on  82  degrees of freedom
##   (35 observations deleted due to missingness)
## AIC: 145.28
##
## Number of Fisher Scoring iterations: 16
##
##
## $'HS graduate or GED'
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -5.385542 257.283445  -0.021  0.98330
## income_cat.L     -0.435599   0.315318  -1.381  0.16714
## income_cat.Q     -0.330887   0.351117  -0.942  0.34600
## ins_classif.L   -10.768890 793.984993  -0.014  0.98918
## ins_classif.Q    -1.728969 965.923981  -0.002  0.99857
## ins_classif.C    -2.410218 682.171999  -0.004  0.99718
## ins_classif^4   -10.092959 668.595490  -0.015  0.98796
## ins_classif^5    -7.328697 515.503116  -0.014  0.98866
## ins_classif^6     4.801493 456.737711   0.011  0.99161
## ins_classif^7    10.706246 869.868008   0.012  0.99018
## race_6cat.L       0.085016   0.752337   0.113  0.91003
## race_6cat.Q       0.033614   0.615301   0.055  0.95643
## race_6cat.C       0.177545   0.645207   0.275  0.78318
## race_6cat^4       0.807011   0.590058   1.368  0.17141
## race_6cat^5       0.414997   0.399870   1.038  0.29935
## sex.L             0.093517   0.238780   0.392  0.69532
## age               0.026384   0.019538   1.350  0.17689
## marital_status.L -0.533525   0.417378  -1.278  0.20115
```

```
## marital_status.Q  -0.190774    0.339915  -0.561   0.57463
## alc_heavyTRUE      -0.246401    0.539072  -0.457   0.64761
## smoking_hx.L       -0.873289    0.322559  -2.707   0.00678 **
## smoking_hx.Q        0.001011    0.289597   0.003   0.99721
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 280.25  on 233  degrees of freedom
## Residual deviance: 247.47  on 212  degrees of freedom
##   (61 observations deleted due to missingness)
## AIC: 291.47
##
## Number of Fisher Scoring iterations: 14
##
##
## $'Some college or AA degree'
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.39930    0.84955  -0.470  0.63835
## income_cat.L     0.28340    0.23771   1.192  0.23317
## income_cat.Q     0.08717    0.29265   0.298  0.76580
## ins_classif.L    0.35960    0.83814   0.429  0.66789
## ins_classif.Q    0.13595    0.75862   0.179  0.85778
## ins_classif.C    0.04305    0.70566   0.061  0.95135
## ins_classif^4    0.98043    0.64830   1.512  0.13046
## ins_classif^5    0.49811    0.56911   0.875  0.38145
## race_6cat.L      1.66038    0.60968   2.723  0.00646 **
## race_6cat.Q      0.39148    0.53114   0.737  0.46109
## race_6cat.C      0.24117    0.53271   0.453  0.65075
## race_6cat^4     -0.73199    0.48262  -1.517  0.12935
## race_6cat^5     -0.48366    0.31380  -1.541  0.12325
## sex.L            0.04252    0.19221   0.221  0.82494
## age              0.02013    0.01305   1.543  0.12294
## marital_status.L -0.42990   0.35432  -1.213  0.22502
## marital_status.Q  0.04120   0.27862   0.148  0.88245
## alc_heavyTRUE     0.08325   0.42510   0.196  0.84475
## smoking_hx.L     -0.31894    0.27474  -1.161  0.24569
## smoking_hx.Q     -0.42390    0.23810  -1.780  0.07502 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 418.47  on 327  degrees of freedom
## Residual deviance: 378.26  on 308  degrees of freedom
##   (59 observations deleted due to missingness)
```

```
## AIC: 418.26
##
## Number of Fisher Scoring iterations: 4
##
##
## $`College graduate or above`
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + marital_status + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.18324  440.34508   0.000  0.99967
## income_cat.L       0.44623    0.42515   1.050  0.29391
## income_cat.Q      -0.20462    0.54988  -0.372  0.70981
## ins_classif.L     -5.54383  788.81665  -0.007  0.99439
## ins_classif.Q     -5.13578  866.00469  -0.006  0.99527
## ins_classif.C     11.71802 1472.36302   0.008  0.99365
## ins_classif^4      7.57584  534.62210   0.014  0.98869
## ins_classif^5     -7.81589 1271.19347  -0.006  0.99509
## ins_classif^6     -5.58280 1902.07158  -0.003  0.99766
## ins_classif^7    -21.08215 1341.21734  -0.016  0.98746
## race_6cat.L        0.21408    0.75356   0.284  0.77635
## race_6cat.Q        0.87742    0.67795   1.294  0.19559
## race_6cat.C        0.49115    0.55806   0.880  0.37880
## race_6cat^4       -0.39226    0.43825  -0.895  0.37075
## race_6cat^5       -0.57940    0.33884  -1.710  0.08727 .
## sex.L              0.01530    0.25875   0.059  0.95284
## age                0.04691    0.01753   2.675  0.00746 **
## marital_status.L   0.67045    0.42830   1.565  0.11750
## marital_status.Q   0.65782    0.39046   1.685  0.09204 .
## alc_heavyTRUE      0.16752    0.73797   0.227  0.82043
## smoking_hx.L       0.07604    0.40653   0.187  0.85162
## smoking_hx.Q      -0.09886    0.35160  -0.281  0.77858
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 304.24  on 241  degrees of freedom
## Residual deviance: 260.69  on 220  degrees of freedom
##   (39 observations deleted due to missingness)
## AIC: 304.69
##
## Number of Fisher Scoring iterations: 15
```

The stratified analysis by education level for adherence to cholesterol medication reveals varying influences of demographic and socioeconomic factors across different educational groups. Here's an interpretation for each group:

1. **Less than 9th grade**: In this group, age was the only variable with a trend towards significance ($p = 0.116$), suggesting a potential increase in adherence with age. However, none of the variables reached

statistical significance, indicating that factors like income, insurance classification, race, sex, marital status, alcohol, and tobacco use did not have a strong, detectable influence on medication adherence in this subgroup.

2. **9th-11th grade**: Here, income (linear term) showed a trend towards significance (p = 0.0835), and age was significantly associated with adherence (p = 0.0353). This suggests that individuals in higher income categories and older ages may be more likely to adhere to medication in this educational group.

3. **High School graduate or GED**: Tobacco use was significantly associated with non-adherence (p = 0.00678), indicating that smokers in this educational group are less likely to adhere to medication. Other factors, including income, insurance, race, sex, marital status, and alcohol use, did not show significant associations.

4. **Some college or AA degree**: Race (linear term) was significantly associated with adherence (p = 0.00646), indicating that racial factors play a role in medication adherence in this group. Other factors like income, insurance, sex, age, marital status, alcohol, and tobacco use did not show a strong association.

5. **College graduate or above**: Age (p = 0.00746) and race (cubic term, p = 0.08727) were significantly associated with adherence, with older individuals and certain racial groups being more likely to adhere to medication. Other variables, including income, insurance, sex, marital status, alcohol, and tobacco use, did not exhibit a significant influence on adherence.

Overall, these results indicate that the factors influencing medication adherence can vary considerably across different education levels. While age consistently appears as a significant or near-significant factor across most groups, the influence of other variables like income, insurance status, race, and health behaviors (smoking and alcohol use) varies. This highlights the complexity of medication adherence as a behavior influenced by a combination of demographic, socioeconomic, and lifestyle factors.

```
lapply(stratified_by_marital, summary)
```

```
## $`Married/Living with Partner`
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.976999  62.076631  -0.064   0.9489
## income_cat.L    0.330274   0.205513   1.607   0.1080
## income_cat.Q   -0.676253   0.269746  -2.507   0.0122 *
## ins_classif.L  -2.477011  38.321069  -0.065   0.9485
## ins_classif.Q   5.441157 191.562328   0.028   0.9773
## ins_classif.C   2.617223  91.689933   0.029   0.9772
## ins_classif^4  -5.741339 180.072070  -0.032   0.9746
## ins_classif^5  -5.663646 159.389228  -0.036   0.9717
## ins_classif^6   4.827395 152.813206   0.032   0.9748
## ins_classif^7   9.903806 296.678477   0.033   0.9734
## race_6cat.L     0.703318   0.444521   1.582   0.1136
## race_6cat.Q     0.430099   0.398504   1.079   0.2805
## race_6cat.C     0.429644   0.337281   1.274   0.2027
## race_6cat^4    -0.164172   0.285099  -0.576   0.5647
## race_6cat^5     0.127339   0.221589   0.575   0.5655
```

```
## sex.L             0.096385   0.156547   0.616   0.5381
## age               0.044213   0.010115   4.371 1.24e-05 ***
## educ_level.L     -0.032267   0.354794  -0.091   0.9275
## educ_level.Q      0.086386   0.316720   0.273   0.7850
## educ_level.C     -0.084653   0.287102  -0.295   0.7681
## educ_level^4     -0.007736   0.247479  -0.031   0.9751
## alc_heavyTRUE     0.654406   0.373007   1.754   0.0794 .
## smoking_hx.L     -0.444670   0.229077  -1.941   0.0522 .
## smoking_hx.Q     -0.213868   0.198237  -1.079   0.2807
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 699.44  on 562  degrees of freedom
## Residual deviance: 614.28  on 539  degrees of freedom
##   (112 observations deleted due to missingness)
## AIC: 662.28
##
## Number of Fisher Scoring iterations: 13
##
##
## $`Widowed/Divorced/Separated`
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    4.17319  276.35710   0.015    0.988
## income_cat.L  -0.08578    0.25465  -0.337    0.736
## income_cat.Q   0.29644    0.28781   1.030    0.303
## ins_classif.L -3.02900  318.47300  -0.010    0.992
## ins_classif.Q -5.28940  551.61013  -0.010    0.992
## ins_classif.C  6.24209  687.97867   0.009    0.993
## ins_classif^4  1.24421  135.79896   0.009    0.993
## ins_classif^5 -7.82780  919.34995  -0.009    0.993
## ins_classif^6  7.83775  831.58319   0.009    0.992
## race_6cat.L    6.84469  291.95221   0.023    0.981
## race_6cat.Q   -0.95873   88.84065  -0.011    0.991
## race_6cat.C   -7.55933  424.81534  -0.018    0.986
## race_6cat^4   -8.76591  461.61556  -0.019    0.985
## race_6cat^5   -4.89918  256.45314  -0.019    0.985
## sex.L          0.18743    0.20477   0.915    0.360
## age            0.02698    0.01647   1.638    0.101
## educ_level.L  -0.36263    0.53957  -0.672    0.502
## educ_level.Q   0.02476    0.45656   0.054    0.957
## educ_level.C   0.11478    0.37627   0.305    0.760
## educ_level^4   0.25379    0.29942   0.848    0.397
## alc_heavyTRUE -0.46605    0.50104  -0.930    0.352
## smoking_hx.L  -0.09962    0.27039  -0.368    0.713
## smoking_hx.Q  -0.17525    0.24057  -0.728    0.466
```

```
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 368.98  on 305  degrees of freedom
## Residual deviance: 345.72  on 283  degrees of freedom
##   (74 observations deleted due to missingness)
## AIC: 391.72
## 
## Number of Fisher Scoring iterations: 15
## 
## 
## $'Never Married'
## 
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + alc_heavy + smoking_hx, family = binomial(link = "logit"),
##     data = data)
## 
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.44531 1616.80862   0.003   0.9973
## income_cat.L     0.93066    0.66768   1.394   0.1634
## income_cat.Q     1.53044    0.80189   1.909   0.0563 .
## ins_classif.L    0.31540 4280.84267   0.000   0.9999
## ins_classif.Q   23.54994 4390.45651   0.005   0.9957
## ins_classif.C   26.69967 3807.74158   0.007   0.9944
## ins_classif^4  -18.20044 3595.71993  -0.005   0.9960
## ins_classif^5  -13.83235 3934.90205  -0.004   0.9972
## race_6cat.L     23.76624 2957.22379   0.008   0.9936
## race_6cat.Q      7.58431 2287.39157   0.003   0.9974
## race_6cat.C    -11.33032 2863.04550  -0.004   0.9968
## race_6cat^4      3.74684 2744.86526   0.001   0.9989
## race_6cat^5    -11.41020 1485.75422  -0.008   0.9939
## sex.L           -0.72902    0.51481  -1.416   0.1567
## age              0.01610    0.02999   0.537   0.5915
## educ_level.L   -10.88090 1753.02278  -0.006   0.9950
## educ_level.Q     8.83680 1481.57468   0.006   0.9952
## educ_level.C    -4.50150  876.51168  -0.005   0.9959
## educ_level^4     2.18238  331.29070   0.007   0.9947
## alc_heavyTRUE    1.25634    0.97570   1.288   0.1979
## smoking_hx.L    -0.78486    0.54508  -1.440   0.1499
## smoking_hx.Q    -0.65956    0.57620  -1.145   0.2523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 120.352  on 87  degrees of freedom
## Residual deviance:  79.409  on 66  degrees of freedom
##   (25 observations deleted due to missingness)
## AIC: 123.41
## 
## Number of Fisher Scoring iterations: 17
```

The stratified analysis by marital status for adherence to cholesterol medication demonstrates that the influence of various factors differs across marital status categories. Here's an interpretation for each group:

1. **Married/Living with Partner**: In this group, age was a significant predictor (p = 1.24e-05), indicating higher adherence with increasing age. Additionally, the quadratic term of income category showed significance (p = 0.0122), suggesting a non-linear relationship between income and adherence. Alcohol use (p = 0.0794) and smoking history (linear term, p = 0.0522) showed trends towards significance, indicating possible influences on adherence.

2. **Widowed/Divorced/Separated**: None of the variables reached statistical significance in this group. This suggests that factors like income, insurance classification, race, sex, age, education level, alcohol, and tobacco use did not have a strong, detectable influence on medication adherence among individuals who are widowed, divorced, or separated.

3. **Never Married**: In this subgroup, the quadratic term of income category showed a trend towards significance (p = 0.0563), hinting at a possible complex relationship between income levels and medication adherence. Other factors, including insurance classification, race, sex, age, education level, alcohol, and tobacco use, did not exhibit a significant influence on adherence.

Overall, these results indicate that marital status can influence how different socioeconomic and demographic factors affect medication adherence. Age consistently appears as a significant predictor in the married/living with partner group, while income shows a complex relationship in the same group and a trend towards significance in the never married group. This highlights the importance of considering marital status when examining factors influencing medication adherence.

```
lapply(stratified_by_alcohol, summary)
```

```
## $`FALSE`
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + marital_status + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.293439  88.983568  -0.015 0.988403
## income_cat.L      0.144029   0.152559   0.944 0.345123
## income_cat.Q     -0.139270   0.182063  -0.765 0.444298
## ins_classif.L    -6.842848 172.956696  -0.040 0.968441
## ins_classif.Q     1.975786 188.828104   0.010 0.991652
## ins_classif.C     9.548265 294.364511   0.032 0.974124
## ins_classif^4    -6.537953 183.386886  -0.036 0.971561
## ins_classif^5    -9.134607 240.000175  -0.038 0.969639
## ins_classif^6     5.839661 306.783565   0.019 0.984813
## ins_classif^7     1.163631 382.761815   0.003 0.997574
## ins_classif^8    11.151314 293.271398   0.038 0.969669
## race_6cat.L       0.847042   0.408311   2.075 0.038033 *
## race_6cat.Q       0.617478   0.364629   1.693 0.090371 .
## race_6cat.C       0.396610   0.302599   1.311 0.189966
## race_6cat^4      -0.115840   0.245099  -0.473 0.636480
## race_6cat^5      -0.151130   0.168630  -0.896 0.370134
## sex.L             0.027096   0.117243   0.231 0.817226
## age               0.027638   0.008071   3.424 0.000616 ***
```

```
## educ_level.L        0.006611    0.278401    0.024 0.981055
## educ_level.Q       -0.088567    0.244700   -0.362 0.717396
## educ_level.C        0.105252    0.212929    0.494 0.621090
## educ_level^4        0.103093    0.181874    0.567 0.570824
## marital_status.L   -0.051080    0.201875   -0.253 0.800247
## marital_status.Q    0.022509    0.168103    0.134 0.893481
## smoking_hx.L        -0.332033    0.165523   -2.006 0.044860 *
## smoking_hx.Q        -0.239962    0.148575   -1.615 0.106290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1081.6  on 870  degrees of freedom
## Residual deviance: 1005.1  on 845  degrees of freedom
##   (66 observations deleted due to missingness)
## AIC: 1057.1
##
## Number of Fisher Scoring iterations: 13
##
##
## $`TRUE`
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + marital_status + smoking_hx, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      6.648e-01  5.656e+02   0.001   0.9991
## income_cat.L     1.777e-01  7.189e-01   0.247   0.8048
## income_cat.Q    -1.283e+00  9.544e-01  -1.344   0.1788
## ins_classif.L    3.236e+00  1.672e+03   0.002   0.9985
## ins_classif.Q    7.920e+00  1.335e+03   0.006   0.9953
## ins_classif.C    1.416e+01  1.539e+03   0.009   0.9927
## ins_classif^4   -4.712e+00  1.434e+03  -0.003   0.9974
## ins_classif^5    5.706e+00  7.708e+02   0.007   0.9941
## race_6cat.L      1.769e+00  1.054e+00   1.678   0.0933 .
## race_6cat.Q      4.130e-01  8.759e-01   0.472   0.6373
## race_6cat.C     -7.684e-01  8.153e-01  -0.943   0.3459
## race_6cat^4      2.591e-01  6.847e-01   0.378   0.7051
## sex.L            3.406e-01  5.064e-01   0.673   0.5013
## age              1.048e-01  3.482e-02   3.011   0.0026 **
## educ_level.L    -5.451e-01  1.099e+00  -0.496   0.6200
## educ_level.Q     1.083e+00  1.019e+00   1.062   0.2881
## educ_level.C    -5.698e-03  8.078e-01  -0.007   0.9944
## educ_level^4     2.569e-02  6.818e-01   0.038   0.9699
## marital_status.L -1.087e+00  7.125e-01  -1.526   0.1269
## marital_status.Q  1.403e+00  6.669e-01   2.104   0.0354 *
## smoking_hx.L     -1.161e+00  6.304e-01  -1.841   0.0656 .
## smoking_hx.Q     -3.514e-01  6.275e-01  -0.560   0.5755
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 112.434  on 85  degrees of freedom
## Residual deviance:  76.298  on 64  degrees of freedom
##   (20 observations deleted due to missingness)
## AIC: 120.3
##
## Number of Fisher Scoring iterations: 15
```

The stratified analysis by alcohol use for adherence to cholesterol medication yields distinct insights for individuals who do not use alcohol (`FALSE`) compared to those who do (`TRUE`).

1. **For Non-Alcohol Users (`FALSE`)**:

   - Age is a significant predictor (p = 0.000616), indicating higher adherence with increasing age.
   - Race, specifically the linear term of race category (p = 0.038033), is significant, suggesting certain racial/ethnic groups might have different adherence patterns.
   - Smoking history, particularly the linear term (p = 0.044860), is significant, hinting at a potential relationship between smoking behavior and medication adherence.
   - Other factors, such as income, insurance classification, sex, education level, and marital status, did not show significant associations with medication adherence.

2. **For Alcohol Users (`TRUE`)**:

   - Age again emerges as a significant factor (p = 0.0026), reinforcing the trend seen in non-alcohol users.
   - Marital status, particularly the quadratic term, is significant (p = 0.0354), suggesting that marital status might play a different role in medication adherence for alcohol users.
   - Race, sex, education level, and other factors did not show significant associations.
   - Smoking history showed a trend towards significance (p = 0.0656 for the linear term), indicating a potential relationship between smoking and adherence in this subgroup.

Overall, these results highlight that alcohol use can modify the relationship between various demographic and socioeconomic factors and medication adherence. Age consistently appears as a significant predictor in both groups, while race and smoking history show significance in the non-alcohol user group. This suggests that interventions to improve medication adherence may need to be tailored differently for individuals based on their alcohol use status.

```
lapply(stratified_by_tobacco, summary)
```

```
## $Never
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + marital_status + alc_heavy, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.425e+00  3.998e+02   0.004  0.99716
## income_cat.L     1.941e-01  2.187e-01   0.887  0.37489
## income_cat.Q    -3.238e-01  2.682e-01  -1.207  0.22740
## ins_classif.L   -1.324e+01  7.886e+02  -0.017  0.98660
```

```
## ins_classif.Q    -1.294e+00  9.055e+02  -0.001  0.99886
## ins_classif.C     1.426e+01  1.225e+03   0.012  0.99071
## ins_classif^4    -8.766e+00  9.318e+02  -0.009  0.99249
## ins_classif^5    -2.830e+00  1.199e+03  -0.002  0.99812
## ins_classif^6    -5.775e+00  1.283e+03  -0.005  0.99641
## ins_classif^7     4.564e+00  1.634e+03   0.003  0.99777
## ins_classif^8     2.112e+01  1.394e+03   0.015  0.98791
## race_6cat.L        9.406e-01  5.524e-01   1.703  0.08862 .
## race_6cat.Q        5.691e-01  4.934e-01   1.153  0.24872
## race_6cat.C        2.869e-01  3.963e-01   0.724  0.46916
## race_6cat^4       -1.684e-01  3.151e-01  -0.534  0.59311
## race_6cat^5       -9.306e-02  2.327e-01  -0.400  0.68915
## sex.L              9.909e-03  1.608e-01   0.062  0.95085
## age                3.900e-02  1.071e-02   3.643  0.00027 ***
## educ_level.L      -4.902e-02  3.843e-01  -0.128  0.89850
## educ_level.Q      -3.015e-01  3.321e-01  -0.908  0.36400
## educ_level.C       1.894e-02  3.167e-01   0.060  0.95231
## educ_level^4       3.248e-01  2.726e-01   1.191  0.23349
## marital_status.L  -4.924e-02  2.806e-01  -0.176  0.86068
## marital_status.Q   1.492e-01  2.475e-01   0.603  0.54665
## alc_heavyTRUE      3.880e-01  4.861e-01   0.798  0.42473
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 608.36  on 491   degrees of freedom
## Residual deviance: 550.36  on 467   degrees of freedom
##   (104 observations deleted due to missingness)
## AIC: 600.36
##
## Number of Fisher Scoring iterations: 15
##
##
## $Past
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + marital_status + alc_heavy, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.010488 222.515462   0.000    1.000
## income_cat.L      0.033101   0.269094   0.123    0.902
## income_cat.Q      0.044064   0.304425   0.145    0.885
## ins_classif.L    -8.247706 412.014432  -0.020    0.984
## ins_classif.Q     0.227639 412.014396   0.001    1.000
## ins_classif.C    13.214935 766.904983   0.017    0.986
## ins_classif^4     0.910472 215.167940   0.004    0.997
## ins_classif^5   -11.243990 647.542285  -0.017    0.986
## ins_classif^6    -1.638750 986.020586  -0.002    0.999
## ins_classif^7   -11.369106 638.103548  -0.018    0.986
## race_6cat.L       0.688003   0.590827   1.164    0.244
```

```
## race_6cat.Q          0.014445    0.525475    0.027     0.978
## race_6cat.C         -0.077697    0.507248   -0.153     0.878
## race_6cat^4          0.345478    0.458101    0.754     0.451
## race_6cat^5         -0.449758    0.317620   -1.416     0.157
## sex.L                0.039077    0.214018    0.183     0.855
## age                  0.008151    0.015326    0.532     0.595
## educ_level.L        -0.206653    0.481401   -0.429     0.668
## educ_level.Q         0.563690    0.421351    1.338     0.181
## educ_level.C        -0.190815    0.366490   -0.521     0.603
## educ_level^4        -0.029788    0.308711   -0.096     0.923
## marital_status.L     0.032698    0.402683    0.081     0.935
## marital_status.Q     0.193109    0.301468    0.641     0.522
## alc_heavyTRUE        0.463864    0.505491    0.918     0.359
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 369.96  on 315  degrees of freedom
## Residual deviance: 340.30  on 292  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 388.3
##
## Number of Fisher Scoring iterations: 14
##
##
## $Current
##
## Call:
## glm(formula = adherence ~ income_cat + ins_classif + race_6cat +
##     sex + age + educ_level + marital_status + alc_heavy, family = binomial(link = "logit"),
##     data = data)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -3.21238    1.44837  -2.218   0.0266 *
## income_cat.L     0.48709    0.34688   1.404   0.1603
## income_cat.Q    -0.28511    0.43781  -0.651   0.5149
## ins_classif.L    0.12511    1.06445   0.118   0.9064
## ins_classif.Q    0.85237    0.91480   0.932   0.3515
## ins_classif.C    0.08516    0.99814   0.085   0.9320
## ins_classif^4    0.73981    1.00171   0.739   0.4602
## ins_classif^5   -1.15318    0.74400  -1.550   0.1211
## race_6cat.L      1.36085    1.16558   1.168   0.2430
## race_6cat.Q      1.18774    1.03580   1.147   0.2515
## race_6cat.C      2.10933    0.96919   2.176   0.0295 *
## race_6cat^4     -0.31776    0.77601  -0.409   0.6822
## race_6cat^5      0.46580    0.48349   0.963   0.3353
## sex.L           -0.11898    0.30124  -0.395   0.6929
## age              0.05765    0.02160   2.669   0.0076 **
## educ_level.L     1.12236    1.16038   0.967   0.3334
## educ_level.Q    -0.57608    1.00930  -0.571   0.5682
## educ_level.C     0.88211    0.67716   1.303   0.1927
## educ_level^4    -0.54101    0.44534  -1.215   0.2244
## marital_status.L -0.28031   0.41063  -0.683   0.4948
## marital_status.Q -0.29069   0.35844  -0.811   0.4174
```

```
## alc_heavyTRUE      0.02783    0.47633    0.058    0.9534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 204.13  on 148  degrees of freedom
## Residual deviance: 169.39  on 127  degrees of freedom
##   (45 observations deleted due to missingness)
## AIC: 213.39
##
## Number of Fisher Scoring iterations: 5
```

The stratified analysis by tobacco use (Never, Past, Current) for adherence to cholesterol medication provides varied results across different groups:

1. **Never Users**:

   - Age is the only significant predictor (p = 0.00027), indicating higher adherence with increasing age.
   - Other variables, including income, insurance classification, race, sex, education level, marital status, and alcohol use, did not show significant associations with adherence.

2. **Past Users**:

   - No variables showed statistical significance in predicting medication adherence.
   - The model includes income, insurance classification, race, sex, age, education level, marital status, and alcohol use, but none demonstrated a clear impact on adherence.

3. **Current Users**:

   - Age and race emerge as significant predictors. Age is significant (p = 0.0076), reinforcing the trend seen in never users.
   - The linear term of race (race_6cat.C) is significant (p = 0.0295), suggesting that certain racial/ethnic groups among current tobacco users might have different adherence patterns.
   - Other factors, including income, insurance classification, sex, education level, marital status, and alcohol use, did not show significant associations with adherence.

These results illustrate that the factors influencing medication adherence may vary based on tobacco use history. Age consistently appears as a significant factor across never and current users. The impact of race is notably different in current users compared to never and past users. This suggests that interventions to improve medication adherence could benefit from considering an individual's tobacco use history.

Stratified Analysis Interpretation

The stratified analyses for adherence to cholesterol medication across various demographic and behavioral groups revealed nuanced insights. Age consistently emerged as a significant factor, particularly among non-tobacco users and current tobacco users, indicating higher adherence with increasing age. Marital status showed significance in certain strata, hinting at the influence of social support systems on adherence. Notably, the influence of race varied across different tobacco usage groups, suggesting diverse adherence patterns among racial groups, especially among current tobacco users. Income, insurance status, and other socioeconomic factors, while included in the models, did not consistently show a significant impact across different strata. This highlights the complexity of medication adherence behavior, which appears to be influenced by a combination of demographic, socioeconomic, and personal behavior factors.

Summary:

The comprehensive study on adherence to cholesterol medication encompassed multiple analytical approaches, revealing intricate patterns influenced by a blend of demographic, socioeconomic, and behavioral factors:

1. **Descriptive Analysis**: This provided a basic understanding of the study population, highlighting key characteristics such as age distribution, gender, income levels, education, marital status, and tobacco and alcohol use. These factors set the foundation for more detailed analyses.

2. **Bivariate Analysis**: This analysis revealed significant associations between medication adherence and various factors. Age emerged as a key determinant, with older individuals showing higher adherence. Marital status also appeared influential, suggesting the role of social support in medication compliance. However, income, insurance, and race/ethnicity associations with adherence were less clear, indicating a need for more nuanced analysis.

3. **Multivariate Analysis**: The logistic regression model integrated various predictors, providing a more comprehensive view. Age maintained its significance, reaffirming its role in adherence. Gender, income, and insurance classifications showed some influence, but their impact was less pronounced than age. Behavioral factors like tobacco use also emerged as significant, pointing to broader health-conscious behaviors among non-users.

4. **Stratified Analysis**: This approach delved deeper into specific subgroups, uncovering distinct patterns within each. Age consistently showed significance across various strata, reinforcing its overall importance. Marital status and race displayed varying levels of influence depending on the stratification factor, such as tobacco use, suggesting complex interactions between these variables and adherence behavior.

Overall, the study findings underscore the multifaceted nature of medication adherence. Age consistently stands out as a crucial factor, with the support system provided by marital status and the lifestyle choices indicated by tobacco use also playing significant roles. The varying influence of socioeconomic factors across different analyses suggests that these elements interact with other variables in complex ways, influencing adherence. These insights are critical for designing targeted interventions and policies to improve medication adherence and address health disparities in cholesterol management.