Child Marriage Phenomenon In Vietnam

Holly Duong

2024-02-29

# load packages  
library(readr)  
library(dplyr)  
library(ggplot2)  
library(gridExtra)  
library(corrplot)  
library(plotly)  
library(reshape2)  
library(car)  
library(kableExtra)  
library(broom)  
library(knitr)

### Setting up Data

# import dataset  
female <- read\_csv("/Users/hollyduong/Desktop/DA 401/ChildMarriageInVietnam/Data/wm.csv")

# get a glimpse of dataset  
head(female)

## # A tibble: 6 × 458  
## HH1 HH2 LN WM1 WM2 WM3 WMINT WM4 WM5 WM6D WM6M WM6Y WM8  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 2 3 1 2 3 92 91 92 19 11 2020 2  
## 2 1 4 2 1 4 2 92 91 92 18 11 2020 1  
## 3 1 9 4 1 9 4 92 91 92 18 11 2020 1  
## 4 1 10 4 1 10 4 92 91 92 18 11 2020 2  
## 5 1 11 4 1 11 4 92 91 92 19 11 2020 2  
## 6 1 11 5 1 11 5 92 91 92 18 11 2020 2  
## # ℹ 445 more variables: WM9 <dbl>, WM17 <dbl>, WM7H <dbl>, WM7M <dbl>,  
## # WM10H <dbl>, WM10M <dbl>, WM11 <dbl>, WM12 <dbl>, WM13 <dbl>, WM14 <dbl>,  
## # WM15 <dbl>, WM22 <dbl>, WM23 <dbl>, WM24 <dbl>, WMHINT <dbl>, WMFIN <dbl>,  
## # WB3M <dbl>, WB3Y <dbl>, WB4 <dbl>, WB5 <dbl>, WB6A <dbl>, WB6B <dbl>,  
## # WB7 <dbl>, WB9 <lgl>, WB10A <lgl>, WB10B <lgl>, WB11 <lgl>, WB12A <lgl>,  
## # WB12B <lgl>, WB14 <dbl>, WB15 <dbl>, WB16 <dbl>, WB17 <dbl>, WB18 <dbl>,  
## # WB19A <chr>, WB19B <chr>, WB19C <chr>, WB19D <chr>, WB19E <chr>, …

# Select the specified columns to create a new dataframe  
female\_df <- select(female, WAGEM, HH6, HH7, MSTATUS, welevel, insurance, ethnicity, windex5, CP2, HA1, MT4, MT9, MT11)  
  
# View the first few rows of the new dataframe  
summary(female\_df)

## WAGEM HH6 HH7 MSTATUS welevel   
## Min. :10.00 Min. :1.000 Min. :1.000 Min. :1.000 Min. :0.00   
## 1st Qu.:18.00 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.00   
## Median :20.00 Median :2.000 Median :3.000 Median :1.000 Median :2.00   
## Mean :20.92 Mean :1.683 Mean :3.404 Mean :1.413 Mean :2.46   
## 3rd Qu.:23.00 3rd Qu.:2.000 3rd Qu.:5.000 3rd Qu.:1.000 3rd Qu.:3.00   
## Max. :47.00 Max. :2.000 Max. :6.000 Max. :9.000 Max. :9.00   
## NA's :2026 NA's :11 NA's :11 NA's :11 NA's :524   
## insurance ethnicity windex5 CP2   
## Min. :1.000 Min. :1.000 Min. :0.000 Min. :1.000   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000   
## Median :1.000 Median :1.000 Median :2.000 Median :1.000   
## Mean :1.135 Mean :2.035 Mean :2.494 Mean :1.431   
## 3rd Qu.:1.000 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:2.000   
## Max. :9.000 Max. :6.000 Max. :5.000 Max. :9.000   
## NA's :524 NA's :864   
## HA1 MT4 MT9 MT11   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.00   
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.00   
## Median :1.000 Median :2.000 Median :1.000 Median :1.00   
## Mean :1.215 Mean :1.664 Mean :1.365 Mean :1.12   
## 3rd Qu.:1.000 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:1.00   
## Max. :9.000 Max. :9.000 Max. :9.000 Max. :9.00   
## NA's :524 NA's :524 NA's :2496 NA's :524

# Rename the columns  
female\_df <- female\_df %>%  
 rename(  
 age\_first\_marriage = WAGEM,  
 area = HH6,  
 region = HH7,  
 marital\_status = MSTATUS,  
 education\_level = welevel,  
 health\_insurance = insurance,  
 ethnicity = ethnicity,  
 wealth\_index = windex5,  
# contraceptive\_decision\_maker = CP5, #R  
 current\_contraceptive\_use = CP2,  
# age\_first\_sexual\_intercourse = SB1, #R  
# currently\_married = MA1,  
# partner\_age = MA2, #R  
 awareness\_hiv\_aids = HA1,  
 used\_computer\_tablet = MT4,  
 used\_internet = MT9,  
 owns\_mobile\_phone = MT11  
)

# Summarize missing values by column  
summarize\_missing <- sapply(female\_df, function(x) sum(is.na(x)))  
print(summarize\_missing)

## age\_first\_marriage area region   
## 2026 11 11   
## marital\_status education\_level health\_insurance   
## 11 524 524   
## ethnicity wealth\_index current\_contraceptive\_use   
## 0 0 864   
## awareness\_hiv\_aids used\_computer\_tablet used\_internet   
## 524 524 2496   
## owns\_mobile\_phone   
## 524

# Recode the value 9 to NA for specified variables  
female\_df <- female\_df %>%  
 mutate(  
 current\_contraceptive\_use = na\_if(current\_contraceptive\_use, 9),  
 used\_internet = na\_if(used\_internet, 9),  
 health\_insurance = na\_if(health\_insurance, 9),  
 education\_level = na\_if(education\_level, 9),  
 awareness\_hiv\_aids = na\_if(awareness\_hiv\_aids, 9),  
 used\_computer\_tablet = na\_if(used\_computer\_tablet, 9),  
 owns\_mobile\_phone = na\_if(owns\_mobile\_phone, 9),  
 marital\_status = na\_if(marital\_status, 9)  
 )  
  
# Recode the value 6 to NA for 'ethnicity'  
female\_df$ethnicity <- na\_if(female\_df$ethnicity, 6)

# Recoding variables from 1 (Yes) and 2 (No) to 1 (Yes) and 0 (No)  
female\_df$health\_insurance <- ifelse(female\_df$health\_insurance == 2, 0, female\_df$health\_insurance)  
female\_df$current\_contraceptive\_use <- ifelse(female\_df$current\_contraceptive\_use == 2, 0, female\_df$current\_contraceptive\_use)  
female\_df$awareness\_hiv\_aids <- ifelse(female\_df$awareness\_hiv\_aids == 2, 0, female\_df$awareness\_hiv\_aids)  
female\_df$used\_computer\_tablet <- ifelse(female\_df$used\_computer\_tablet == 2, 0, female\_df$used\_computer\_tablet)  
female\_df$owns\_mobile\_phone <- ifelse(female\_df$owns\_mobile\_phone == 2, 0, female\_df$owns\_mobile\_phone)  
female\_df$used\_internet <- ifelse(female\_df$used\_internet == 2, 0, female\_df$used\_internet)

# View the changes to ensure NA substitution has been correctly applied  
summary(female\_df)

## age\_first\_marriage area region marital\_status   
## Min. :10.00 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:18.00 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000   
## Median :20.00 Median :2.000 Median :3.000 Median :1.000   
## Mean :20.92 Mean :1.683 Mean :3.404 Mean :1.408   
## 3rd Qu.:23.00 3rd Qu.:2.000 3rd Qu.:5.000 3rd Qu.:1.000   
## Max. :47.00 Max. :2.000 Max. :6.000 Max. :3.000   
## NA's :2026 NA's :11 NA's :11 NA's :18   
## education\_level health\_insurance ethnicity wealth\_index   
## Min. :0.00 Min. :0.0000 Min. :1.000 Min. :0.000   
## 1st Qu.:1.00 1st Qu.:1.0000 1st Qu.:1.000 1st Qu.:1.000   
## Median :2.00 Median :1.0000 Median :1.000 Median :2.000   
## Mean :2.46 Mean :0.8659 Mean :1.586 Mean :2.494   
## 3rd Qu.:3.00 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:4.000   
## Max. :5.00 Max. :1.0000 Max. :4.000 Max. :5.000   
## NA's :525 NA's :525 NA's :1148   
## current\_contraceptive\_use awareness\_hiv\_aids used\_computer\_tablet  
## Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.0000   
## Median :1.0000 Median :1.0000 Median :0.0000   
## Mean :0.5842 Mean :0.7975 Mean :0.3412   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000   
## NA's :885 NA's :541 NA's :532   
## used\_internet owns\_mobile\_phone  
## Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:1.0000   
## Median :1.0000 Median :1.0000   
## Mean :0.6394 Mean :0.8831   
## 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000   
## NA's :2501 NA's :529

### Creating New Variable Access to Media

# Combine individual variables for access to the internet, phone, and computer into a single variable  
# This new variable "access\_to\_media" will have a value of 1 if the respondent has access to any of these media sources, and 0 if not  
# This provides a more comprehensive measure of media access  
female\_df <- female\_df %>%  
 mutate(access\_to\_media = ifelse(used\_computer\_tablet == 1 | used\_internet == 1 | owns\_mobile\_phone == 1, 1, 0))  
  
# In later analysis, use access\_to\_media instead of the 3 separate variables

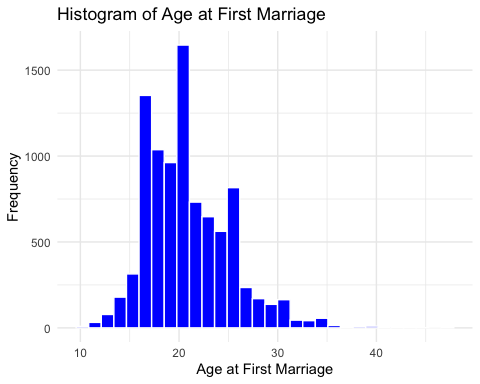
# Exporting female\_df to a CSV file in the current working directory  
#write.csv(female\_df, "female\_df.csv", row.names = FALSE)

### Distributions of Data

# Histogram for 'Age at First Marriage'  
afm\_hist <- ggplot(female\_df, aes(x = age\_first\_marriage)) +  
 geom\_histogram(fill = "blue", color = "white") +  
 theme\_minimal() +  
 ggtitle("Histogram of Age at First Marriage") +  
 xlab("Age at First Marriage") +  
 ylab("Frequency")  
  
print(afm\_hist)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

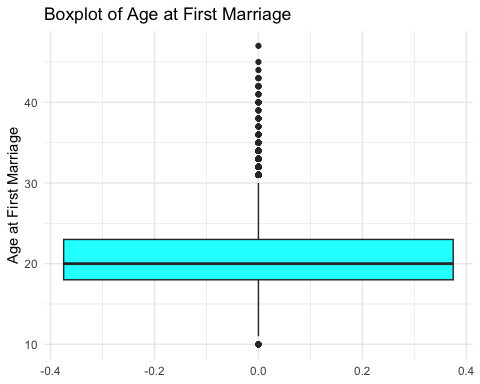
## Warning: Removed 2026 rows containing non-finite outside the scale range  
## (`stat\_bin()`).



# Saving the histogram  
#ggsave("histogram\_age\_first\_marriage.png", plot = afm\_hist, width = 8, height = 6, dpi = 300)

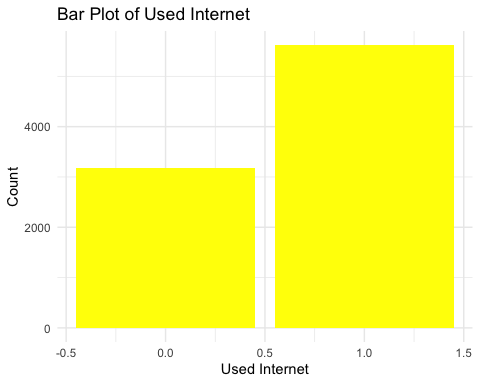
# Plotting a boxplot for 'Age at First Marriage'  
ggplot(female\_df, aes(y = age\_first\_marriage)) +  
 geom\_boxplot(fill = "cyan") +  
 labs(title = "Boxplot of Age at First Marriage", y = "Age at First Marriage") +  
 theme\_minimal()

## Warning: Removed 2026 rows containing non-finite outside the scale range  
## (`stat\_boxplot()`).



# Bar plot for 'used\_internet'  
ggplot(female\_df, aes(x = used\_internet)) +  
 geom\_bar(fill = "yellow") +  
 theme\_minimal() +  
 ggtitle("Bar Plot of Used Internet") +  
 xlab("Used Internet") +  
 ylab("Count")

## Warning: Removed 2501 rows containing non-finite outside the scale range  
## (`stat\_count()`).



### Handling Missing Data

# Before that, let's double check the count of missing values by column  
count\_missing <- sapply(female\_df, function(x) sum(is.na(x)))  
print(count\_missing)

## age\_first\_marriage area region   
## 2026 11 11   
## marital\_status education\_level health\_insurance   
## 18 525 525   
## ethnicity wealth\_index current\_contraceptive\_use   
## 1148 0 885   
## awareness\_hiv\_aids used\_computer\_tablet used\_internet   
## 541 532 2501   
## owns\_mobile\_phone access\_to\_media   
## 529 529

# Make a copy of female\_df for imputation  
imputed\_df <- female\_df

# Handle missing values in 'Age at First Marriage'  
  
# Calculate the median value for 'Age at First Marriage', excluding NA values  
median\_age\_first\_marriage <- median(imputed\_df$age\_first\_marriage, na.rm = TRUE)  
  
# Impute missing values in 'Age at First Marriage' with the median value  
imputed\_df$age\_first\_marriage[is.na(imputed\_df$age\_first\_marriage)] <- median\_age\_first\_marriage

# Define a function to calculate mode for categorical variables  
getMode <- function(v) {  
 # The mode is the value that appears most frequently in the data  
 uniqv <- unique(na.omit(v)) # Omit NA values and get unique values  
 uniqv[which.max(tabulate(match(v, uniqv)))] # Return the value with the highest frequency  
}

# For 'ethnicity', an ordinal variable with predefined categories, it makes sense to impute missing values with the mode.  
mode\_ethnicity <- getMode(imputed\_df$ethnicity)  
imputed\_df <- mutate(imputed\_df, ethnicity = ifelse(is.na(ethnicity), mode\_ethnicity, ethnicity))  
  
# Binary variables like 'current\_contraceptive\_use', 'health\_insurance','awareness\_hiv\_aids', 'used\_internet', 'used\_computer\_tablet', 'owns\_mobile\_phone', and 'access\_to\_media'  
# should be imputed with the mode since it represents the most frequent category (either 0 or 1).  
  
# Calculate the mode for each binary variable  
mode\_used\_internet <- getMode(imputed\_df$used\_internet)  
mode\_current\_contraceptive\_use <- getMode(imputed\_df$current\_contraceptive\_use)  
mode\_health\_insurance <- getMode(imputed\_df$health\_insurance)  
mode\_awareness\_hiv\_aids <- getMode(imputed\_df$awareness\_hiv\_aids)  
mode\_used\_computer\_tablet <- getMode(imputed\_df$used\_computer\_tablet)  
mode\_owns\_mobile\_phone <- getMode(imputed\_df$owns\_mobile\_phone)  
mode\_access\_to\_media <- getMode(imputed\_df$access\_to\_media)  
  
# Impute missing values for binary variables  
imputed\_df <- mutate(imputed\_df,  
 used\_internet = ifelse(is.na(used\_internet), mode\_used\_internet, used\_internet),  
 current\_contraceptive\_use = ifelse(is.na(current\_contraceptive\_use), mode\_current\_contraceptive\_use, current\_contraceptive\_use),  
 health\_insurance = ifelse(is.na(health\_insurance), mode\_health\_insurance, health\_insurance),  
 awareness\_hiv\_aids = ifelse(is.na(awareness\_hiv\_aids), mode\_awareness\_hiv\_aids, awareness\_hiv\_aids),  
 used\_computer\_tablet = ifelse(is.na(used\_computer\_tablet), mode\_used\_computer\_tablet, used\_computer\_tablet),  
 owns\_mobile\_phone = ifelse(is.na(owns\_mobile\_phone), mode\_owns\_mobile\_phone, owns\_mobile\_phone),  
 access\_to\_media = ifelse(is.na(access\_to\_media), mode\_access\_to\_media, access\_to\_media)  
)  
  
# 'education\_level' is an ordinal variable where the median could be a more suitable measure of central tendency than the mode.  
# However, given the categorical nature of the levels (e.g., "Primary", "Secondary"), using the mode may still be appropriate.  
mode\_education\_level <- getMode(imputed\_df$education\_level)  
imputed\_df <- mutate(imputed\_df, education\_level = ifelse(is.na(education\_level), mode\_education\_level, education\_level))

# Check the resulting dataset to confirm changes  
summary(imputed\_df)

## age\_first\_marriage area region marital\_status   
## Min. :10.00 Min. :1.000 Min. :1.000 Min. :1.000   
## 1st Qu.:18.00 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000   
## Median :20.00 Median :2.000 Median :3.000 Median :1.000   
## Mean :20.76 Mean :1.683 Mean :3.404 Mean :1.408   
## 3rd Qu.:23.00 3rd Qu.:2.000 3rd Qu.:5.000 3rd Qu.:1.000   
## Max. :47.00 Max. :2.000 Max. :6.000 Max. :3.000   
## NA's :11 NA's :11 NA's :18   
## education\_level health\_insurance ethnicity wealth\_index   
## Min. :0.000 Min. :0.0000 Min. :1.000 Min. :0.000   
## 1st Qu.:1.000 1st Qu.:1.0000 1st Qu.:1.000 1st Qu.:1.000   
## Median :2.000 Median :1.0000 Median :1.000 Median :2.000   
## Mean :2.438 Mean :0.8721 Mean :1.527 Mean :2.494   
## 3rd Qu.:3.000 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:4.000   
## Max. :5.000 Max. :1.0000 Max. :4.000 Max. :5.000   
##   
## current\_contraceptive\_use awareness\_hiv\_aids used\_computer\_tablet  
## Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.0000   
## Median :1.0000 Median :1.0000 Median :0.0000   
## Mean :0.6168 Mean :0.8072 Mean :0.3251   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## used\_internet owns\_mobile\_phone access\_to\_media   
## Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:1.0000   
## Median :1.0000 Median :1.0000 Median :1.0000   
## Mean :0.7192 Mean :0.8886 Mean :0.9071   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000   
##

# After imputation, let's check for missing values  
count\_imputation <- sapply(imputed\_df, function(x) sum(is.na(x)))  
print(count\_imputation)

## age\_first\_marriage area region   
## 0 11 11   
## marital\_status education\_level health\_insurance   
## 18 0 0   
## ethnicity wealth\_index current\_contraceptive\_use   
## 0 0 0   
## awareness\_hiv\_aids used\_computer\_tablet used\_internet   
## 0 0 0   
## owns\_mobile\_phone access\_to\_media   
## 0 0

### Visualization Comparing Before vs. After Imputation

# Histogram for 'age\_first\_marriage' before imputation  
afm\_0 <- ggplot(female\_df, aes(x = age\_first\_marriage)) +   
 geom\_histogram(fill = "lightpink", color = "white", bins = 30) +  
 theme\_light() +  
 ggtitle("Age at First Marriage Before Imputation") +  
 xlab("Age at First Marriage") +  
 ylab("Frequency")  
  
# Histogram for 'age\_first\_marriage' after imputation  
afm\_imputed <- ggplot(imputed\_df, aes(x = age\_first\_marriage)) +   
 geom\_histogram(fill = "plum", color = "white", bins = 30) +  
 theme\_light() +  
 ggtitle("Age at First Marriage After Imputation") +  
 xlab("Age at First Marriage") +  
 ylab("Frequency")  
  
# Arrange the two plots side by side  
grid.arrange(afm\_0, afm\_imputed, ncol = 2)

## Warning: Removed 2026 rows containing non-finite outside the scale range  
## (`stat\_bin()`).



# Arrange the two plots side by side and capture the layout as a grob  
combined\_plots <- arrangeGrob(afm\_0, afm\_imputed, ncol = 2)

## Warning: Removed 2026 rows containing non-finite outside the scale range  
## (`stat\_bin()`).

# Now, use ggsave to save the combined plot  
ggsave("combined\_age\_first\_marriage.png", plot = combined\_plots, width = 10, height = 5)

### Creating Binary Variables for Child Marriage Under 18 and 15

# Convert "age at first marriage" into a binary variable to indicate child marriage  
# Child marriage is defined as marriage before the age of 18  
# The new binary variable "child\_marriage" will have a value of 1 if the marriage occurred before age 18, and 0 otherwise  
imputed\_df <- imputed\_df %>%  
 mutate(child\_marriage = ifelse(age\_first\_marriage < 18, 1, 0))

# Create a binary variable for child marriage under 15  
# The new variable "child\_marriage\_u15" will have a value of 1 if the marriage occurred before age 15, and 0 otherwise  
imputed\_df <- imputed\_df %>%  
 mutate(child\_marriage\_u15 = ifelse(age\_first\_marriage < 15, 1, 0))

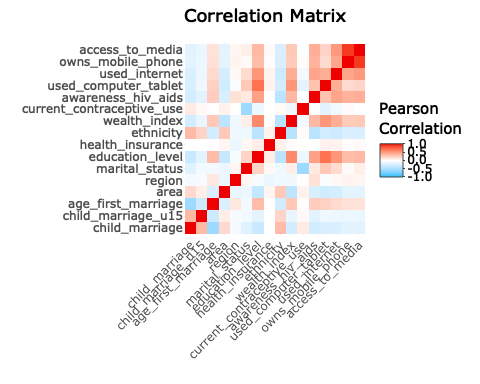
# Move "child\_marriage" and "child\_marriage\_u15" to the front of the dataframe  
imputed\_df <- imputed\_df %>%  
 select(child\_marriage, child\_marriage\_u15, everything())

### Preparing for Regression Analysis

#### Correlation Matrix

# Calculate correlation matrix  
cor\_matrix <- cor(imputed\_df %>% select\_if(is.numeric), use = "complete.obs")  
  
# Melt the correlation matrix  
melted\_cor\_matrix <- melt(cor\_matrix)

# Generate an interactive heatmap  
corr\_matrix <- ggplot(melted\_cor\_matrix, aes(Var1, Var2, fill = value)) +  
 geom\_tile() +  
 scale\_fill\_gradientn(  
 colours = c("deepskyblue", "white", "red2"),  
 values = scales::rescale(c(-1, 0, 1)),  
 limits = c(-1, 1),  
 name="Pearson\nCorrelation"  
 ) +  
 theme\_minimal() +   
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 xlab("") +   
 ylab("") +  
 ggtitle("Correlation Matrix")   
  
# Convert ggplot object to plotly for interactivity  
ggplotly(corr\_matrix)



#### Accessing Multicollinearity

# Model with all predictors  
model\_for\_vif <- lm(child\_marriage ~ area + region + marital\_status + education\_level + health\_insurance + ethnicity + wealth\_index + current\_contraceptive\_use + awareness\_hiv\_aids + access\_to\_media, data = imputed\_df)  
  
# Calculate VIF (A VIF value > 5 indicates high multicollinearity)  
vif\_results <- vif(model\_for\_vif)  
print(vif\_results)

## area region marital\_status   
## 1.276710 1.111376 1.556525   
## education\_level health\_insurance ethnicity   
## 1.931734 1.040051 1.429007   
## wealth\_index current\_contraceptive\_use awareness\_hiv\_aids   
## 1.774619 1.501796 1.556100   
## access\_to\_media   
## 1.251097

#### Converting Categorical and Binary Variables to Factors

# Convert nominal and ordinal variables to factors  
imputed\_df$area <- as.factor(imputed\_df$area)  
imputed\_df$region <- as.factor(imputed\_df$region)  
imputed\_df$marital\_status <- as.factor(imputed\_df$marital\_status)  
imputed\_df$education\_level <- factor(imputed\_df$education\_level, ordered = TRUE)  
imputed\_df$ethnicity <- as.factor(imputed\_df$ethnicity)  
imputed\_df$wealth\_index <- factor(imputed\_df$wealth\_index, ordered = TRUE)  
  
# Binary variables are already in the correct format and can be used as is

### Baseline Logistic Regression Model

baseline\_model <- glm(child\_marriage ~ area + region + marital\_status + education\_level + health\_insurance + ethnicity + wealth\_index + current\_contraceptive\_use + awareness\_hiv\_aids + access\_to\_media, family = binomial(), data = imputed\_df)  
  
# Check the model summary to interpret the coefficients  
summary(baseline\_model)

##   
## Call:  
## glm(formula = child\_marriage ~ area + region + marital\_status +   
## education\_level + health\_insurance + ethnicity + wealth\_index +   
## current\_contraceptive\_use + awareness\_hiv\_aids + access\_to\_media,   
## family = binomial(), data = imputed\_df)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.69204 0.19574 -13.753 < 2e-16 \*\*\*  
## area2 0.23405 0.08675 2.698 0.006972 \*\*   
## region2 0.31846 0.13180 2.416 0.015683 \*   
## region3 0.11824 0.13020 0.908 0.363836   
## region4 0.28828 0.12882 2.238 0.025226 \*   
## region5 -0.01239 0.12758 -0.097 0.922641   
## region6 -0.07003 0.13543 -0.517 0.605094   
## marital\_status2 0.19873 0.12129 1.639 0.101311   
## marital\_status3 -16.69086 132.38178 -0.126 0.899668   
## education\_level.L -2.63955 0.24416 -10.811 < 2e-16 \*\*\*  
## education\_level.Q -0.79978 0.15213 -5.257 1.46e-07 \*\*\*  
## education\_level.C 0.73946 0.28177 2.624 0.008681 \*\*   
## education\_level^4 0.81761 0.29403 2.781 0.005423 \*\*   
## education\_level^5 0.25180 0.17168 1.467 0.142454   
## health\_insurance 0.03453 0.08420 0.410 0.681703   
## ethnicity2 -0.02966 0.10707 -0.277 0.781759   
## ethnicity3 0.27077 0.13057 2.074 0.038108 \*   
## ethnicity4 1.21560 0.10944 11.107 < 2e-16 \*\*\*  
## wealth\_index.L -0.43723 0.13081 -3.343 0.000830 \*\*\*  
## wealth\_index.Q -0.18217 0.11943 -1.525 0.127181   
## wealth\_index.C 0.27357 0.10035 2.726 0.006408 \*\*   
## wealth\_index^4 -0.32427 0.08513 -3.809 0.000139 \*\*\*  
## wealth\_index^5 -0.10307 0.07948 -1.297 0.194712   
## current\_contraceptive\_use 0.09590 0.06954 1.379 0.167884   
## awareness\_hiv\_aids -0.15285 0.07775 -1.966 0.049312 \*   
## access\_to\_media -0.06893 0.08757 -0.787 0.431210   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 10432.2 on 11275 degrees of freedom  
## Residual deviance: 7737.2 on 11250 degrees of freedom  
## (18 observations deleted due to missingness)  
## AIC: 7789.2  
##   
## Number of Fisher Scoring iterations: 17