# Load the packages  
library(readxl)

## Warning: package 'readxl' was built under R version 4.4.1

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.4.1

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.4.1

# Define the file path (replace with your actual file path)  
file\_path <- "C:\\Users\\arthu\\Downloads\\Book1.xlsx"  
  
# Read the Excel file (assuming the sheet name is "Sheet1", modify if different)  
data <- read\_excel(file\_path)  
  
# Display the first few rows of the dataset to understand its structure  
print("First few rows of the dataset:")

## [1] "First few rows of the dataset:"

print(head(data))

## # A tibble: 6 × 14  
## ID `Marital Status` Gender Income Children Education Occupation   
## <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr>   
## 1 12496 Maried Female 40000 1 Bachelors Skilled Manual  
## 2 24107 Maried Male 30000 3 Partial College Clerical   
## 3 14177 Maried Male 80000 5 Partial College Professional   
## 4 24381 Single Male 70000 0 Bachelors Professional   
## 5 25597 Single Male 30000 0 Bachelors Clerical   
## 6 13507 Maried Female 10000 2 Partial College Manual   
## # ℹ 7 more variables: `Home Owner` <chr>, Cars <dbl>, `Commute Distance` <chr>,  
## # Region <chr>, Age <dbl>, `Age Brackets` <chr>, `Purchased Bike` <chr>

# Summary statistics of the dataset  
print("Summary statistics of the dataset:")

## [1] "Summary statistics of the dataset:"

print(summary(data))

## ID Marital Status Gender Income   
## Min. :11000 Length:1000 Length:1000 Min. : 10000   
## 1st Qu.:15291 Class :character Class :character 1st Qu.: 30000   
## Median :19744 Mode :character Mode :character Median : 60000   
## Mean :19966 Mean : 56360   
## 3rd Qu.:24471 3rd Qu.: 70000   
## Max. :29447 Max. :170000   
## Children Education Occupation Home Owner   
## Min. :0.000 Length:1000 Length:1000 Length:1000   
## 1st Qu.:0.000 Class :character Class :character Class :character   
## Median :2.000 Mode :character Mode :character Mode :character   
## Mean :1.898   
## 3rd Qu.:3.000   
## Max. :5.000   
## Cars Commute Distance Region Age   
## Min. :0.000 Length:1000 Length:1000 Min. :25.00   
## 1st Qu.:1.000 Class :character Class :character 1st Qu.:35.00   
## Median :1.000 Mode :character Mode :character Median :43.00   
## Mean :1.442 Mean :44.16   
## 3rd Qu.:2.000 3rd Qu.:52.00   
## Max. :4.000 Max. :89.00   
## Age Brackets Purchased Bike   
## Length:1000 Length:1000   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##

# Basic Analysis  
# Count of Purchased Bike by Gender  
bike\_purchase\_by\_gender <- data %>%  
 group\_by(Gender) %>%  
 summarise(Count = sum(`Purchased Bike` == "Yes"))  
  
print("Count of Purchased Bike by Gender:")

## [1] "Count of Purchased Bike by Gender:"

print(bike\_purchase\_by\_gender)

## # A tibble: 2 × 2  
## Gender Count  
## <chr> <int>  
## 1 Female 239  
## 2 Male 242

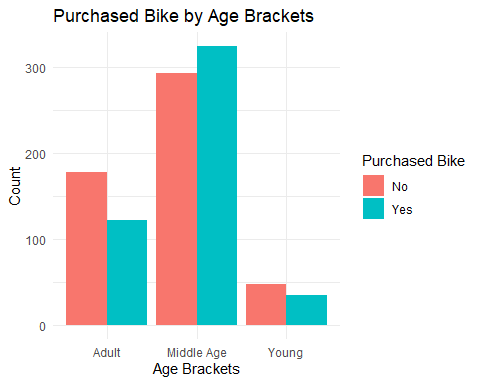
# Average Income by Education Level  
avg\_income\_by\_education <- data %>%  
 group\_by(Education) %>%  
 summarise(Average\_Income = mean(Income, na.rm = TRUE))  
  
print("Average Income by Education Level:")

## [1] "Average Income by Education Level:"

print(avg\_income\_by\_education)

## # A tibble: 5 × 2  
## Education Average\_Income  
## <chr> <dbl>  
## 1 Bachelors 63007.  
## 2 Graduate Degree 66092.  
## 3 High School 47263.  
## 4 Partial College 54717.  
## 5 Partial High School 34474.

# Plot: Purchased Bike by Age Brackets  
ggplot(data, aes(x = `Age Brackets`, fill = `Purchased Bike`)) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Purchased Bike by Age Brackets", x = "Age Brackets", y = "Count") +  
 theme\_minimal()



# Plot: Average Income by Education Level  
ggplot(avg\_income\_by\_education, aes(x = Education, y = Average\_Income)) +  
 geom\_bar(stat = "identity", fill = "steelblue") +  
 labs(title = "Average Income by Education Level", x = "Education Level", y = "Average Income") +  
 theme\_minimal()

