ANA 515 Assignment 2, Loading, Saving, and Describing Data

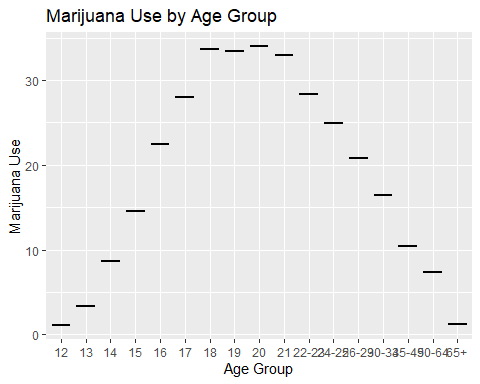
Anuj Modi

2023-06-21

library(ggplot2)  
library(dplyr)  
  
# Read the dataset  
df <- read.csv("drug-use-by-age.csv")

Visualization 1: Boxplot of Marijuana Use by Age Group

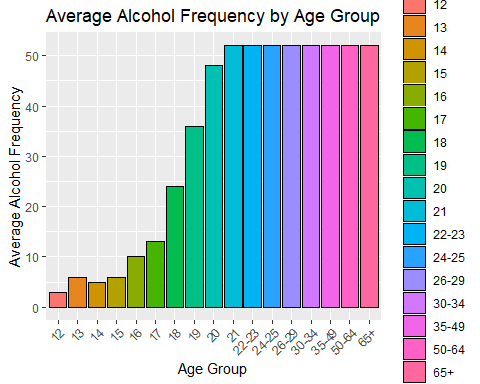
# Create a boxplot of marijuana use by age group  
ggplot(df, aes(x = age, y = marijuana\_use)) +  
 geom\_boxplot(fill = "skyblue", color = "black") +  
 labs(title = "Marijuana Use by Age Group",  
 x = "Age Group",  
 y = "Marijuana Use")



The boxplot allows us to visually compare the distribution of marijuana use across different age groups. Each box represents the interquartile range (IQR) of marijuana use within an age group, with the median marked by a horizontal line. This visualization helps us identify any variations in marijuana use among different age categories

Visualization 2: Bar Plot of Alcohol Frequency by Age Group

# Create a bar plot of alcohol frequency by age group  
df %>%   
 group\_by(age) %>%  
 summarise(mean\_alcohol\_freq = mean(alcohol\_frequency, na.rm = TRUE)) %>%  
 ggplot(aes(x = age, y = mean\_alcohol\_freq, fill = age)) +  
 geom\_bar(stat = "identity", color = "black") +  
 labs(title = "Average Alcohol Frequency by Age Group",  
 x = "Age Group",  
 y = "Average Alcohol Frequency") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



The bar plot displays the average alcohol frequency for each age group. Each bar represents the average alcohol frequency, indicating the typical frequency of alcohol use within each age category. This visualization helps us compare the average alcohol consumption patterns across different age groups.

These visualizations provide initial insights into marijuana use by age group and average alcohol frequency among different age categories in the dataset on drug use by age. They allow us to explore and understand the data in terms of these specific factors.