Demographics

Tolentino, Omisol, and Espadon

2024-05-20

```
# Install the required packages
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("tidyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
# Load the required libraries
library(ggplot2)
library(dplyr)
```

```
##
## 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(tidyr)
```

```
# Read the CSV file
data <- read.csv("/cloud/project/Survey/survey.csv")
```

```
# Remove whitespace from column names
names(data) <- trimws(names(data))
```

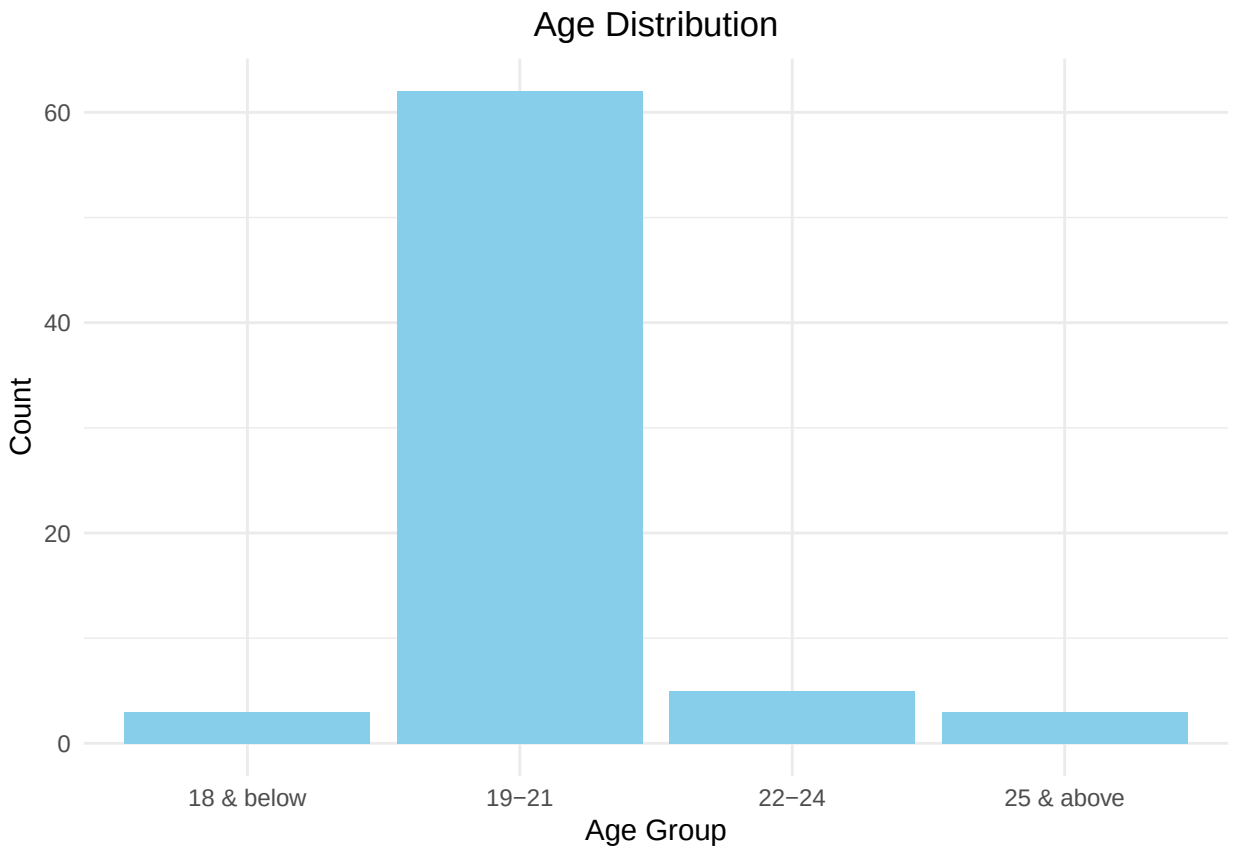
```
# Standardize age groups
data$Age. <- gsub("17|18", "18 & below", data$Age.)
data$Age. <- gsub("19|20|21|19 yrs. Old", "19-21", data$Age.)
data$Age. <- gsub("22|23|24", "22-24", data$Age.)
data$Age. <- gsub("25|26|27|28", "25 & above", data$Age.)
```

```
# Convert columns to appropriate data types
data$Age. <- as.factor(data$Age.)
data$Sex. <- as.factor(data$Sex.)
```

```r
data$Education.level <- as.factor(data$Education.level)
data$How.often.do.you.use.Digital.Payment. <- as.factor(data$How.often.do.you.use.Digital.Payment.)
data$How.long.have.you.been.using.Digital.payment. <- as.factor(data$How.long.have.you.been.using.Digita
```

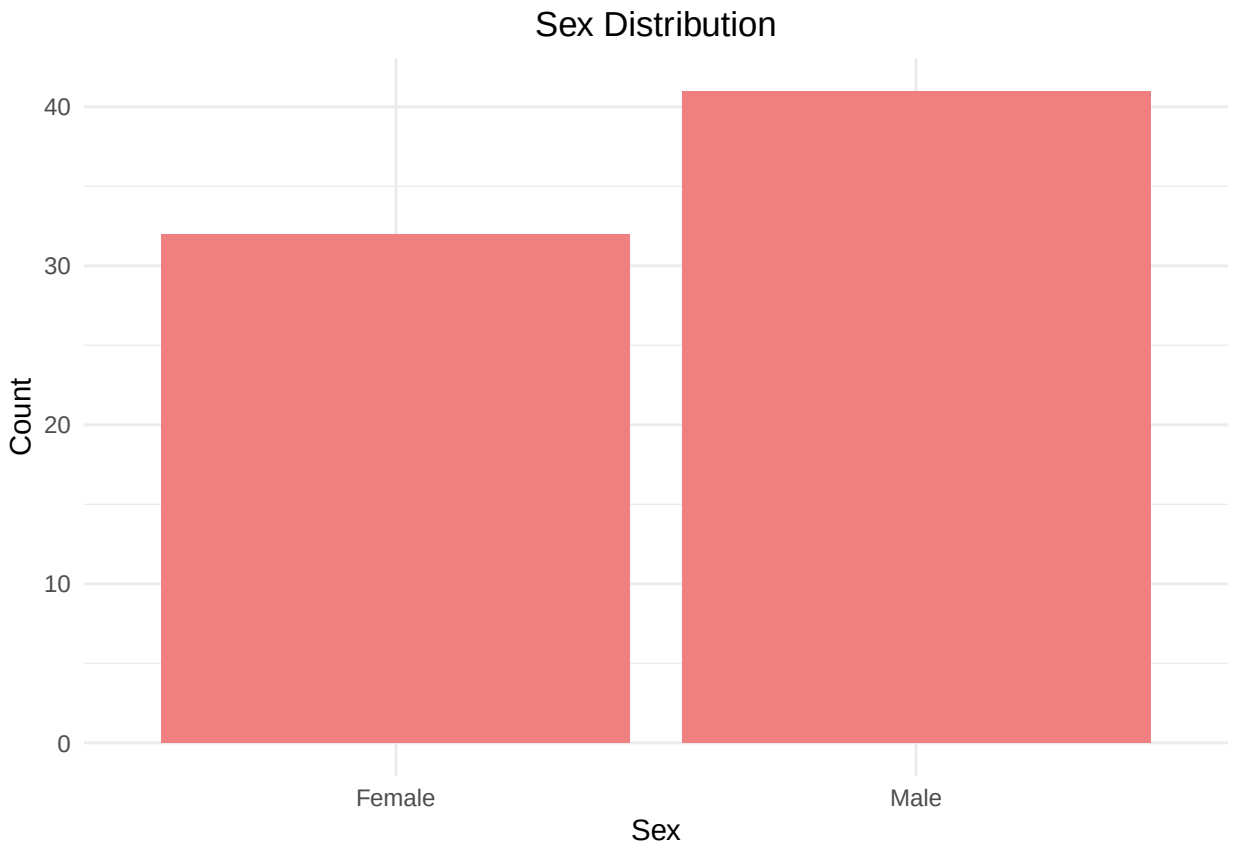
```r
# Remove rows with missing data
cleaned_data <- na.omit(data)

# Remove duplicate rows
cleaned_data <- cleaned_data %>%
  distinct()
```

```r
# Create bar plot for Age distribution
bar_plot_age <- ggplot(cleaned_data, aes(x = Age.)) +
  geom_bar(fill = "skyblue") +
  labs(title = "Age Distribution",
       x = "Age Group",
       y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

bar_plot_age
```



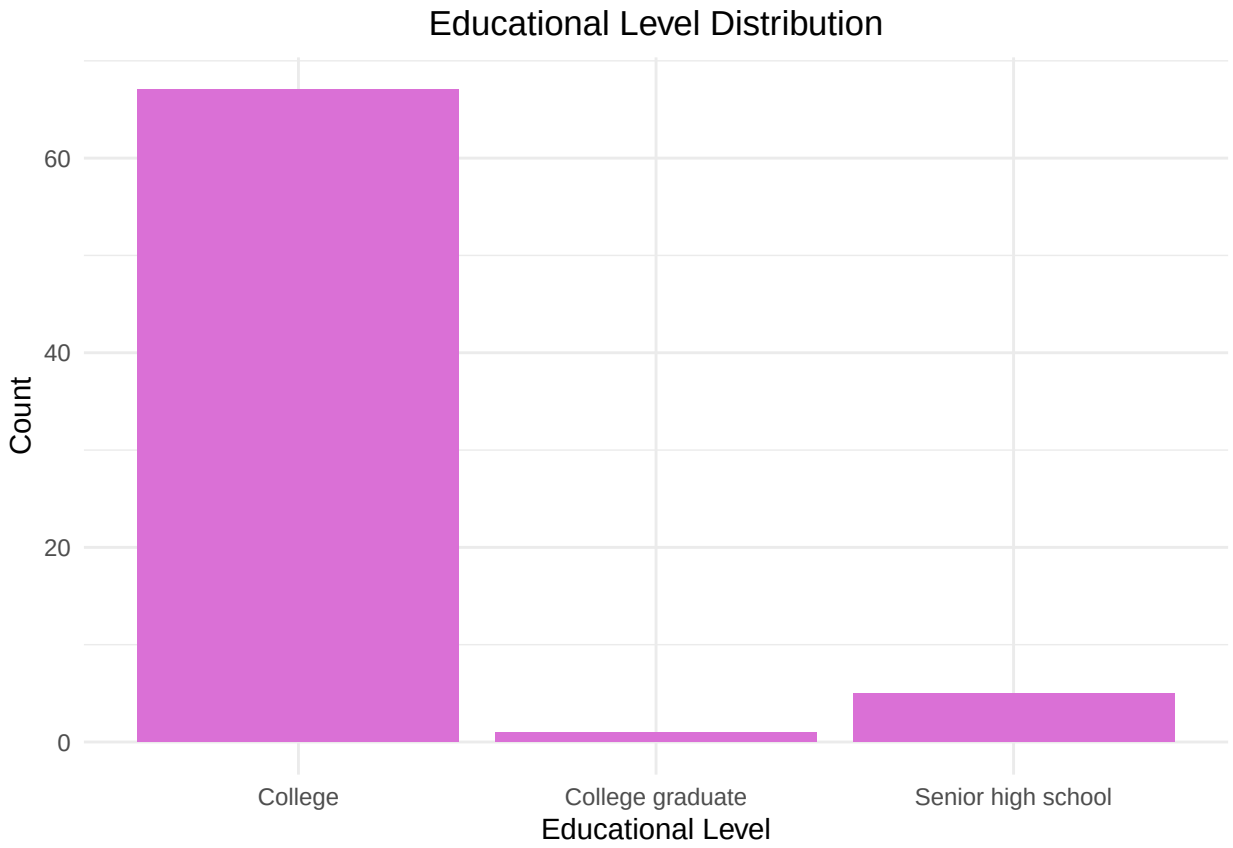

```r
#The bar plot illustrates the distribution of survey respondents across distinct age categories.
#The age ranges are segmented into four groups: "18 & below", "19-21", "22-24", and "25 & above".
#Each bar on the plot represents the frequency of individuals falling within a specific age range.
#It's evident that the "19-21" age range has the highest frequency.
```

```r
# Create bar plot for Sex distribution
bar_plot_sex <- ggplot(cleaned_data, aes(x = Sex.)) +
  geom_bar(fill = "lightcoral") +
  labs(title = "Sex Distribution",
       x = "Sex",
       y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

bar_plot_sex
```



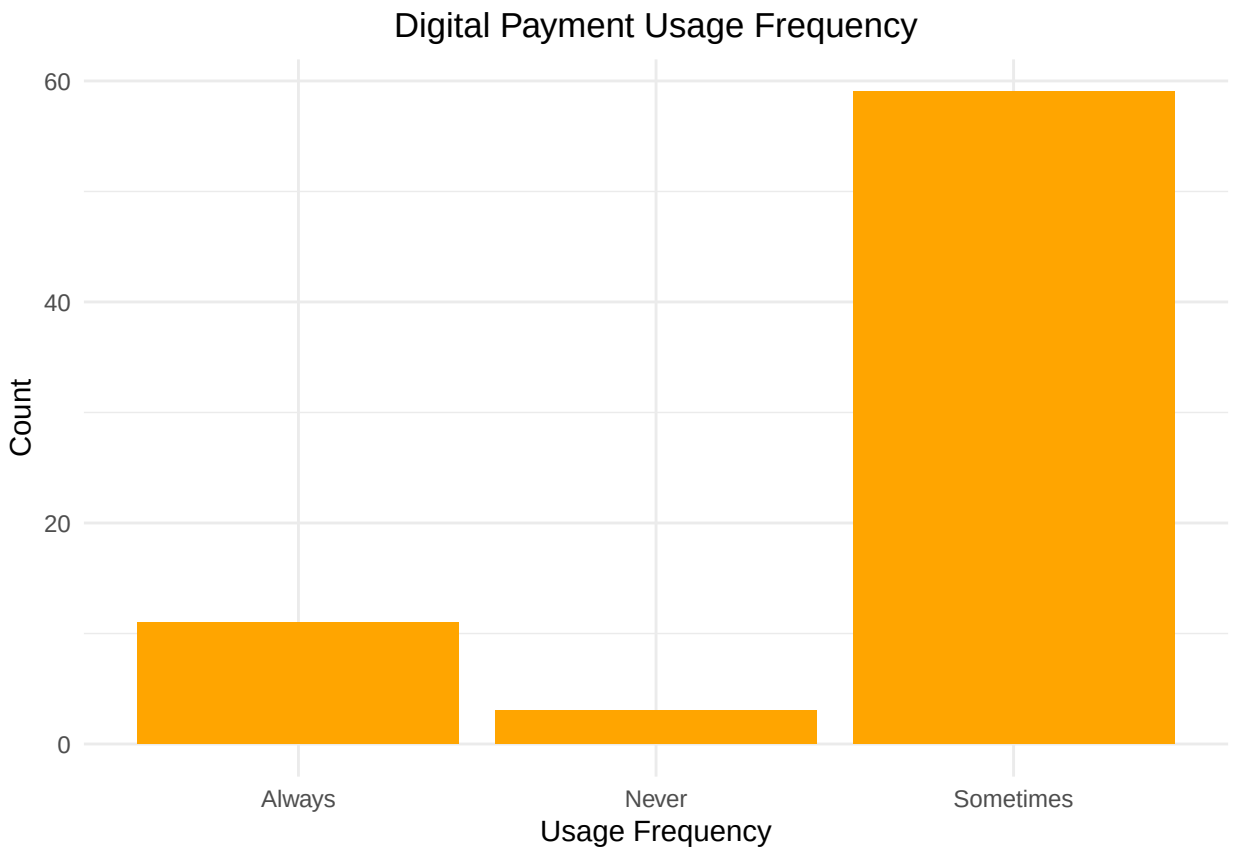

```r
#The bar plotcategorizing individuals into two groups: "Male" and "Female".
#Each bar on the plot represents the frequency of respondents belonging to a specific sex category.
#We can identify which sex category has the highest frequency of respondents which is "Male".
```

```r
# Create bar plot for Education Level distribution
bar_plot_educlevel <- ggplot(cleaned_data, aes(x = Education.level)) +
  geom_bar(fill = "orchid") +
  labs(title = "Educational Level Distribution",
       x = "Educational Level",
       y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

bar_plot_educlevel
```



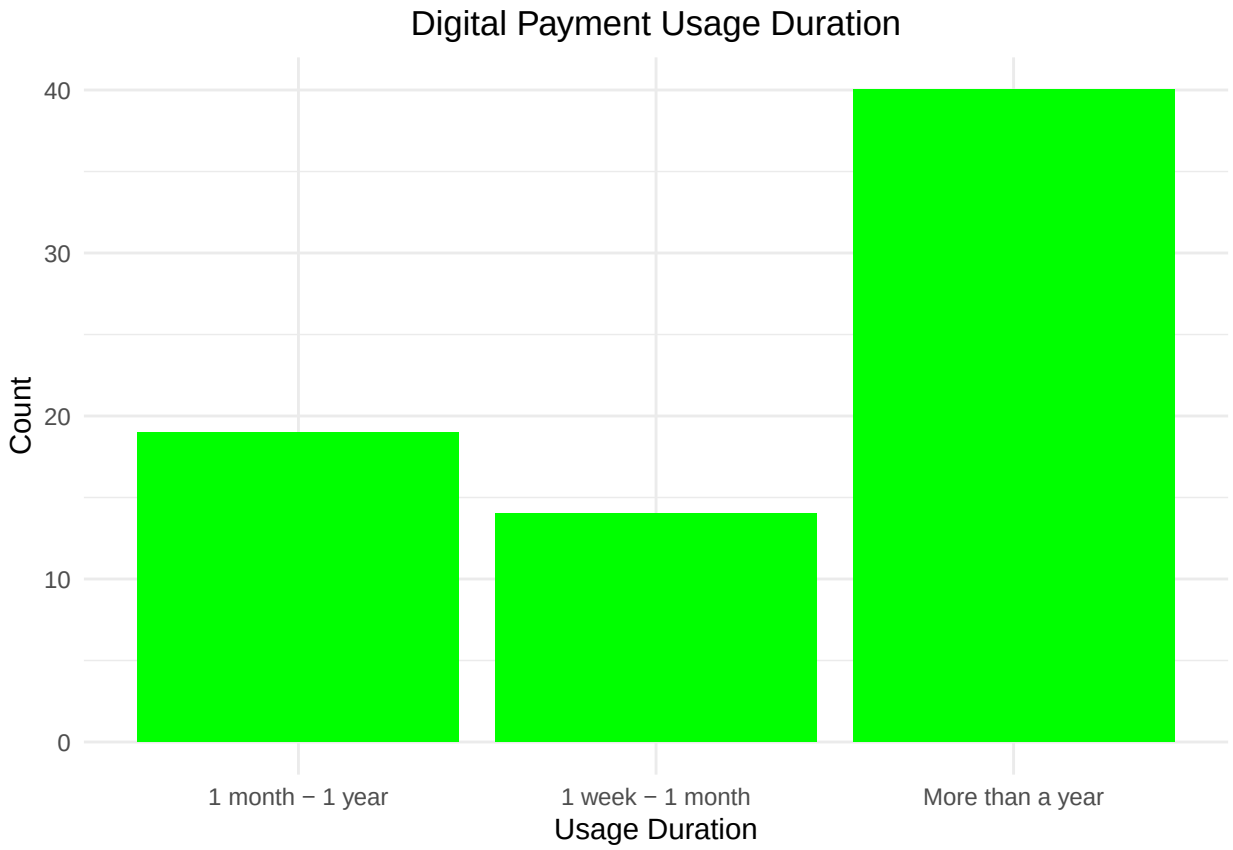

```
#The highest frequency is observed in the "College" category.
#Indicating that the majority of respondents have completed or are currently attending college.
#Conversely, the lowest frequency is found in the "College graduate" category.
#Suggesting that fewer respondents have completed only senior high school education.
```

```
# Create bar plot for Digital Payment Usage Frequency
bar_plot_often <- ggplot(cleaned_data, aes(x = How.often.do.you.use.Digital.Payment.)) +
  geom_bar(fill = "orange") +
  labs(title = "Digital Payment Usage Frequency",
       x = "Usage Frequency",
       y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

bar_plot_often
```




```
#The bar plot depicts the frequency of respondents' usage of digital payment applications.
#Categorizing users into three frequency categories: "Never", "Sometimes", and "Always".
#Each bar represents thefrequency of respondents falling into a particular usage frequency category.
#From the bar plot, the "Sometimes" category has the highest frequency of respondents.
#While the "Never" category has the lowest frequency of respondents.
```

```
# Create bar plot for Digital Payment Usage Duration
bar_plot_long <- ggplot(cleaned_data, aes(x = How.long.have.you.been.using.Digital.payment.)) +
  geom_bar(fill = "green") +
  labs(title = "Digital Payment Usage Duration",
       x = "Usage Duration",
       y = "Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

bar_plot_long
```




```
#The bar plot visualizes the duration of time users have been utilizing digital payment applications.
#Categorizing users into different time intervals
#Such as "1 week - 1 month", "1 month - 1 year", and"More than a year."
#From the bar plot, we find that the category "More than a year" has the highest frequency.
#While the category "1 week - 1 month" has the lowest frequency of respondents.
```