16_262_Tushar_R6-7_PassionPersuitSurvey.R

tushar sakuja

2023-11-28

```
# Importing different libraries
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
##
## 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 4.3.2
```

```
library(ggalluvial)
```

```
## Warning: package 'ggalluvial' was built under R version 4.3.2
```

```
library("rgl")
```

```
## Warning: package 'rgl' was built under R version 4.3.2
```

```
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.3.2
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

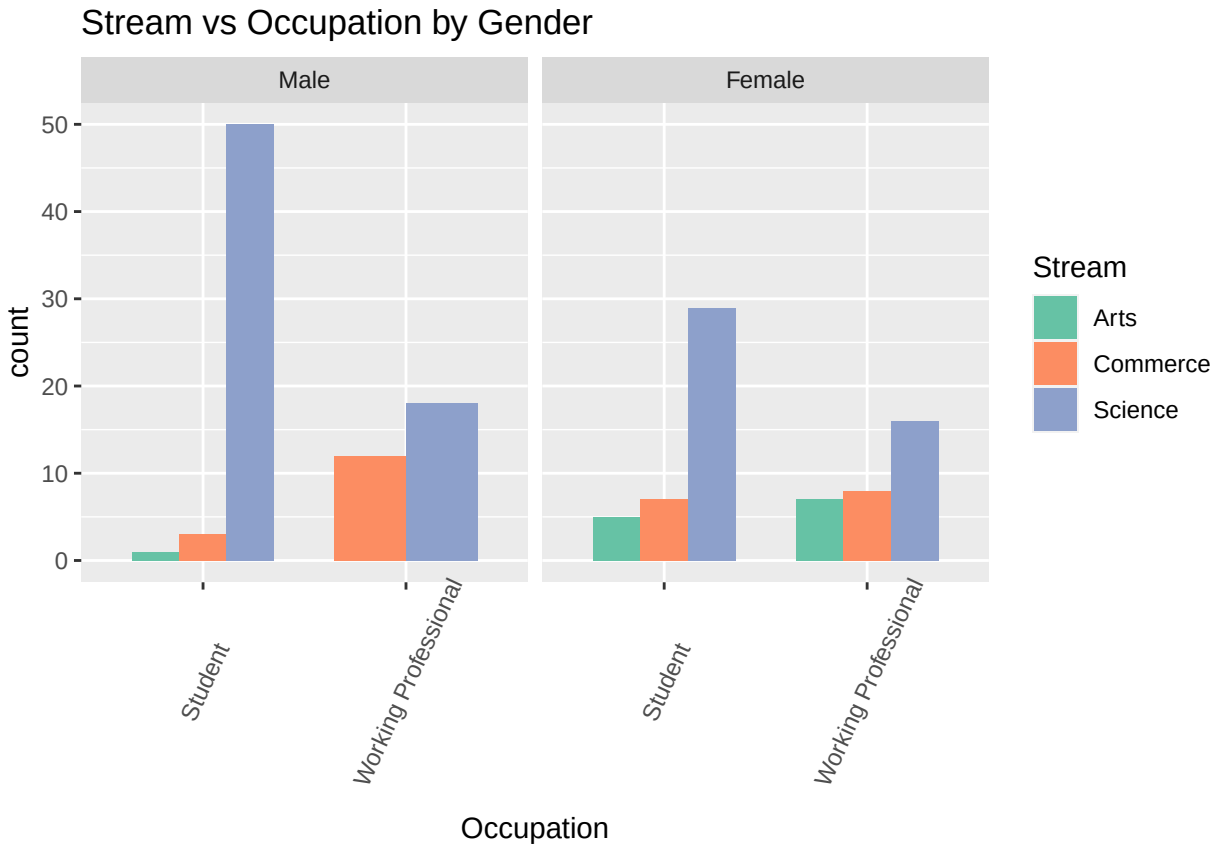
```
library("stringr")
```

```
## Warning: package 'stringr' was built under R version 4.3.2
```

```
# Reading the CSV file and rename the columns
passion <- read.csv("C:/Users/tushar sakuja/Desktop/SEMESTER-2/APPLIED STATS WITH R/Final_R_Project/16_
colnames(passion) <- c("Email", "Name", "Age", "Gender", "Occupation","Stream",
                       "StudiedPassion", "StudyInspiration", "JobPreference",
                       "WorkOnPassion", "CareerGuidance", "GotCareerGuidance",
                       "WhenCareerGuidance", "WorkLifeBalance",
                       "EducationInIndia", "FuturePreference", "WhyIndia",
                       "WhyDecision", "ViewIndia", "WhyAbroad", "WhyDecision",
                       "ConsideredChallenges", "WhatChallenges", "ViewAbroad")
duplicated_columns <- duplicated(names(passion))

# Printing the dataset
head(passion)
```

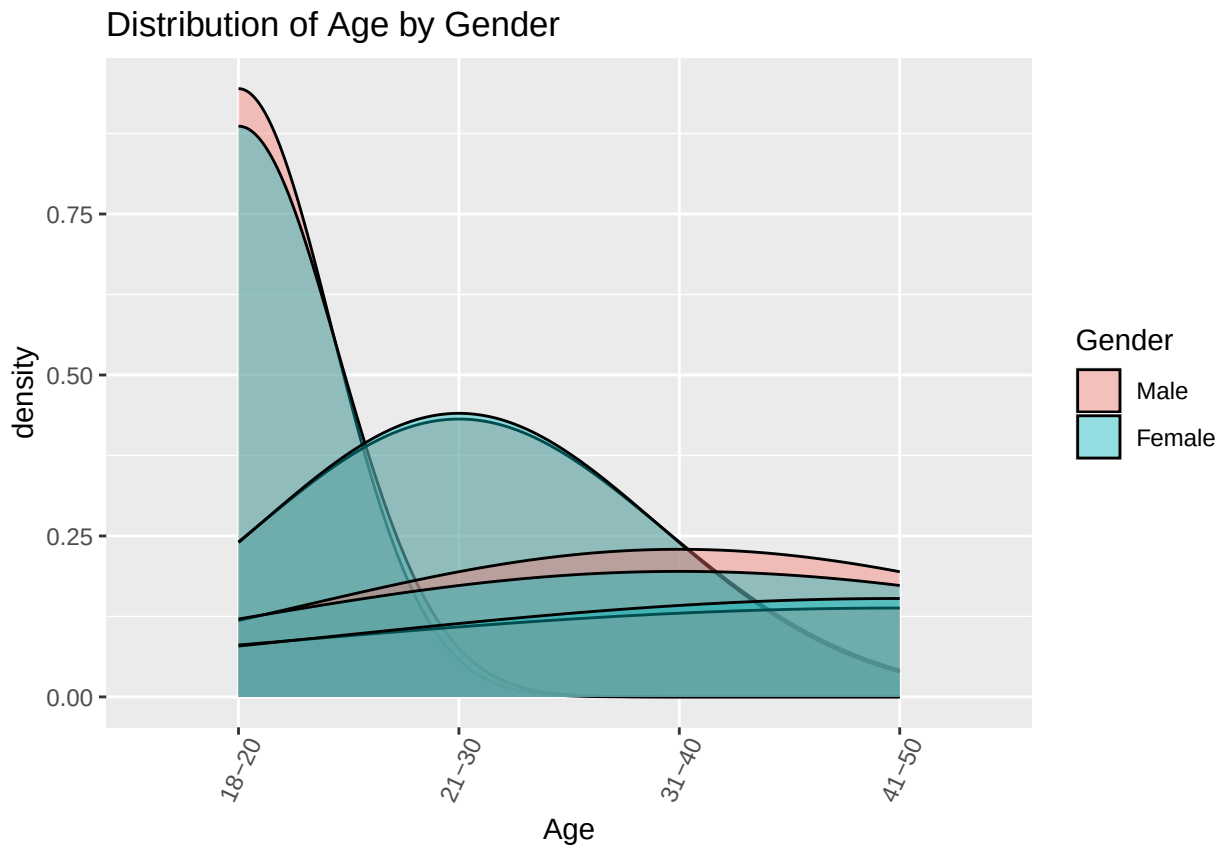
```
##                         Email             Name   Age Gender           Occupation   Stream StudiedPassi
## 1     pyalibasu.12@gmail.com            Pyali 21-30 Female Working Professional Commerce            Y
## 2  shreya.basu.jsr@gmail.com       Shreya Roy 21-30 Female Working Professional  Science
## 3 riyaratnakaran02@gmail.com             Riya 21-30 Female Working Professional  Science            Y
## 4       rkr1236300@gmail.com  Rushikesh Reddy 21-30   Male Working Professional  Science
## 5 shrutinarayana19@gmail.com Shruti Narayana  21-30 Female Working Professional  Science
## 6  shrutisakhuja94@gmail.com  Shruti Sakhuja  21-30 Female Working Professional Commerce            Y
##         JobPreference WorkOnPassion CareerGuidance GotCareerGuidance    WhenCareerGuidance     WorkLi
## 1 Passion related job           Yes            Yes                No          Grade 12/PUC Extremely
## 2   Higher salary job         Maybe            Yes                No          Grade 12/PUC Extremely
## 3 Passion related job           Yes            Yes               Yes Undergraduate Courses Extremely
## 4   Higher salary job            No             No               Yes Undergraduate Courses Extremely
## 5   Higher salary job           Yes            Yes               Yes          Grade 12/PUC  Somewhat
## 6 Passion related job           Yes            Yes               Yes          Grade 12/PUC Extremely
##   EducationInIndia FuturePreference                        WhyIndia          WhyDecision ViewIndia
## 1               No           Abroad                                                              Bet
## 2               No            India To stay close to one's family Personal preferences
## 3              Yes            India To stay close to one's family Personal preferences
## 4               No           Abroad                                                              Bet
## 5              Yes            India To stay close to one's family Personal preferences
## 6              Yes            India         Inclusive and Diverse     Economic factors
##            WhyDecision ConsideredChallenges     WhatChallenges ViewAbroad
## 1 Personal preferences                  Yes Legal requirements
## 2
## 3
## 4     Economic factors                   No Legal requirements
## 5
## 6
```

```r
# Removing the duplicate columns
if (any(duplicated_columns)) {
  for (i in which(duplicated_columns)) {
    new_name <- paste(names(passion)[i], "_2")
    names(passion)[i] <- new_name
  }
}

# Convert Gender to a factor
passion$Gender <- factor(passion$Gender, levels = c("Male", "Female"))

# Plotting bar graph  Stream vs Occupation by Gender
ggplot(passion, aes(x = Occupation, fill = Stream)) +
  geom_bar(position = "dodge", width = 0.7) +
  labs(title = "Stream vs Occupation by Gender") +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6)) +
  facet_wrap(~ Gender) +  # Adding facets for different genders
  scale_fill_brewer(palette = "Set2")
```

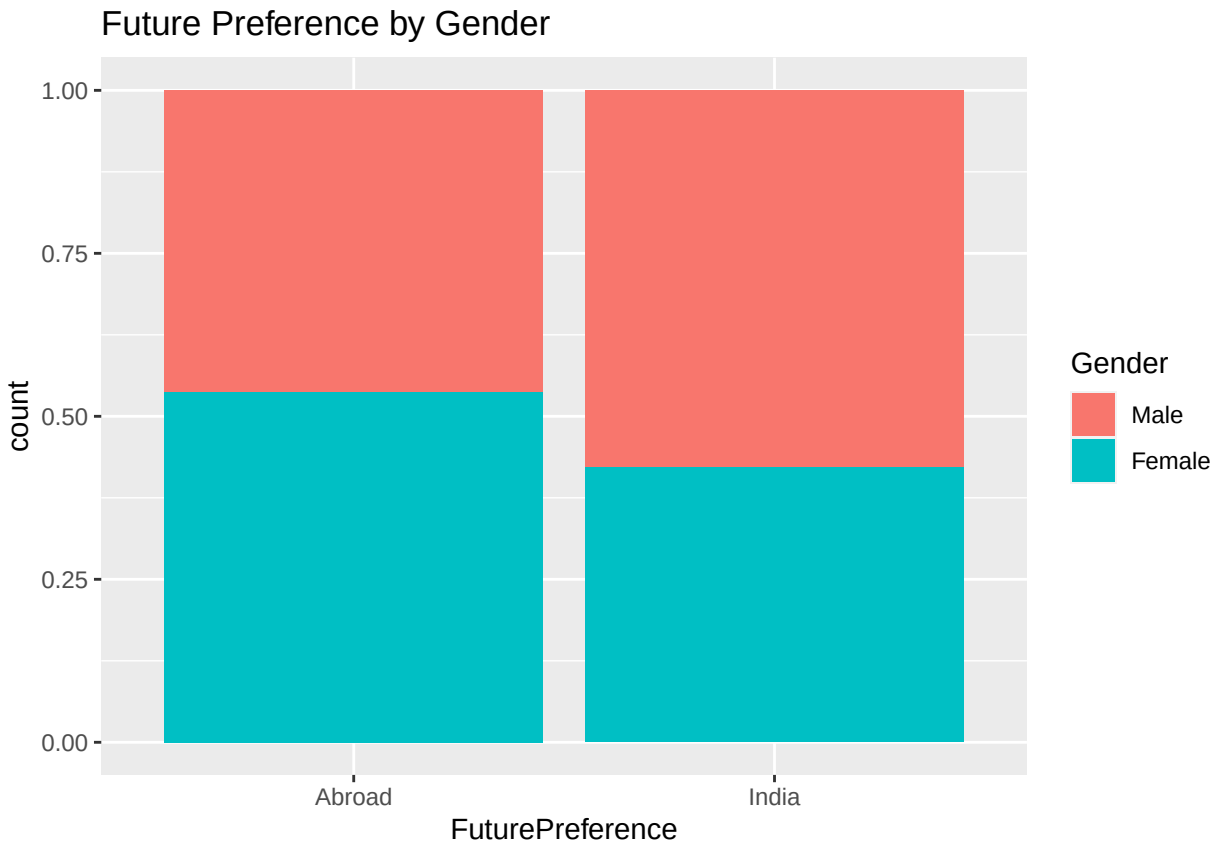



```r
#=========================================================================
# INTERPRETATION :- Educational stream choices vary by occupation, with working
#professionals more likely to pursue streams related to their fields of study.
#Men are more inclined towards Science and Commerce streams, while women favor
#Arts and, Science and Commerce streams as well . Occupation seems to play a
#more significant role than gender in influencing educational preferences.
```

```r
#=============================================================================

# Plotting distribution of age by gender
ggplot(passion, aes(x = Age, fill = Gender)) +
  geom_density(alpha = 0.4) +
  labs(title = "Distribution of Age by Gender") +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6))
```

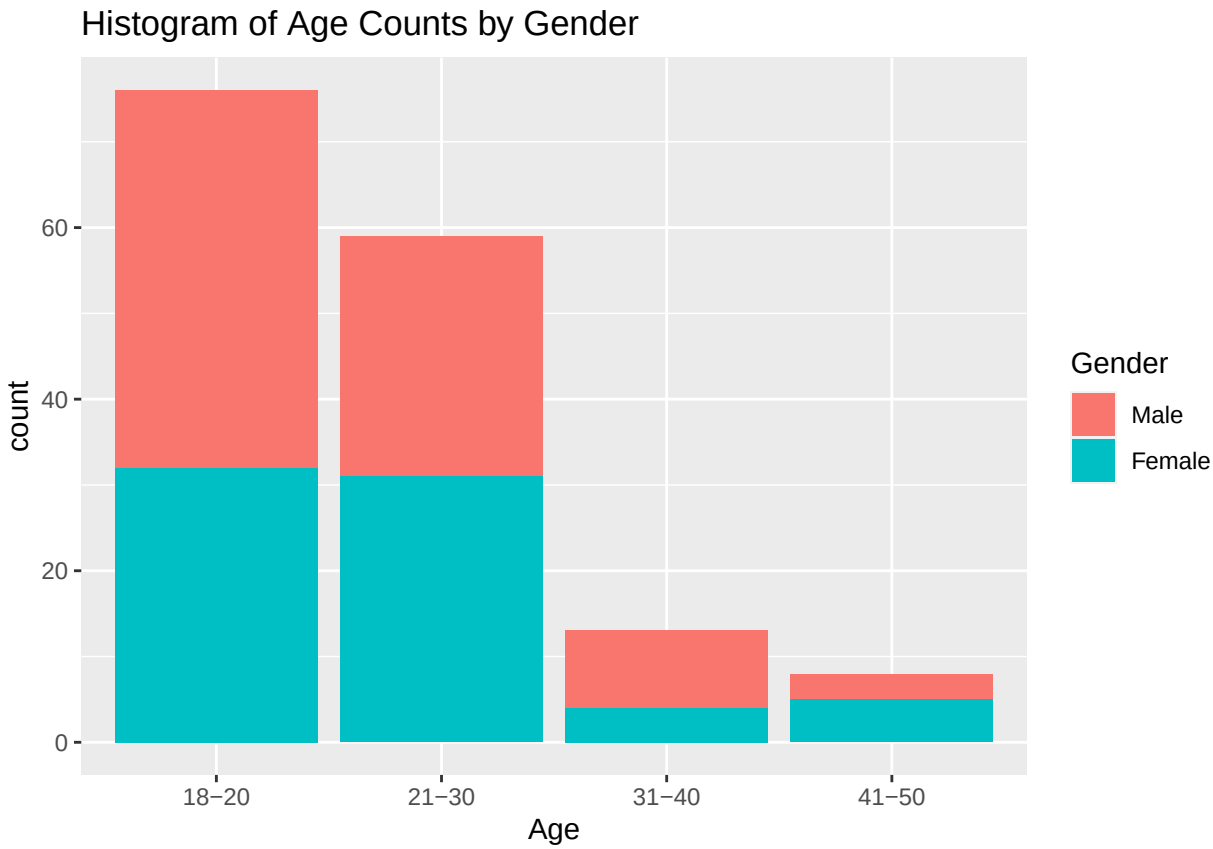



```r
#=============================================================================
# INTERPRETATION :- The graph shows that the median age is higher for males than
#females. This could be due to a number of factors, such as differences in life
#expectancy, educational background, or employment patterns. In the graph the
#age distribution is skewed to the right, meaning that there are more people in
#the older age groups than in the younger age groups.
#=============================================================================

# Plotting bar plot for Future Preference by Gender
ggplot(passion, aes(x = FuturePreference, fill = Gender)) +
  geom_bar(position = "fill") +
  labs(title = "Future Preference by Gender")
```

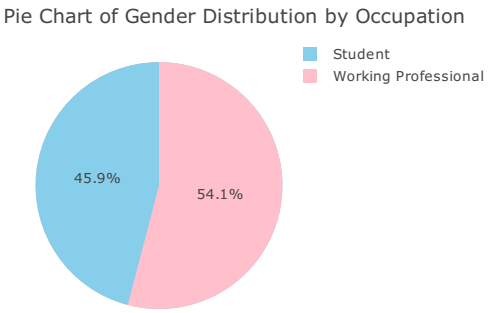



```
#=========================================================================
# INTERPRETATION :- The graph shows that a higher percentage of people prefer to
#work or study in abroad than in India. This trend is more pronounced among men
#than women.
#=========================================================================

# Plotting Histogram of Age Counts by Gender
ggplot(passion, aes(x = Age, fill = Gender)) +
  geom_histogram(stat = "count") +
  labs(title = "Histogram of Age Counts by Gender")
```

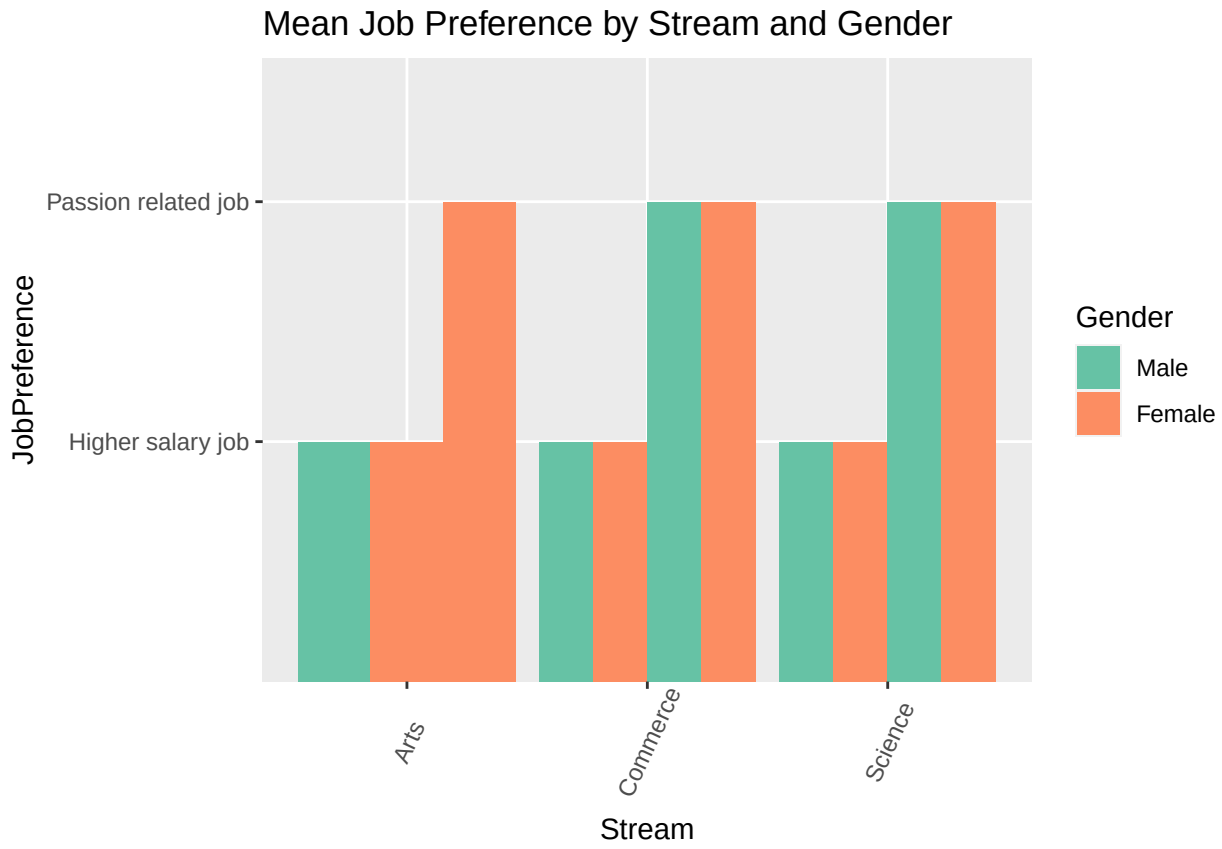
```
## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad
```




```
#=============================================================================
# INTERPRETATION :- The histogram shows the distribution of age by gender, with
#more women than men in the younger age groups and more men than women in the
#older age groups. This suggests that women tend to be younger than men on
#average or women have a higher life expectancy than men.
#=============================================================================

# Calculation of the proportions for the pie chart
gender_occupation <- table(passion$Gender, passion$Occupation)
prop <- prop.table(gender_occupation, margin = 2)

# Creating the pie chart
plot_ly(labels = colnames(prop), values = prop[1, ],
        type = "pie", name = "Male",
        marker = list(colors = c("skyblue", "pink"))) %>%
  add_trace(values = prop[2, ], name = "Female") %>%
  layout(title = "Pie Chart of Gender Distribution by Occupation",
         scene = list(aspectmode = "data"))
```




```
#=============================================================================
# INTERPRETATION :- The gender distribution of the workforce varies depending on
#the occupation, with some occupations being predominantly female and others
#being predominantly male.
#=============================================================================

# Plotting bar graph for Mean Job Preference by Stream and Gender
ggplot(passion, aes(x = Stream, y = JobPreference, fill = Gender)) +
  geom_bar(stat = "summary", fun = "mean", position = "dodge") +
  labs(title = "Mean Job Preference by Stream and Gender") +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6)) +
  scale_fill_brewer(palette = "Set2")
```

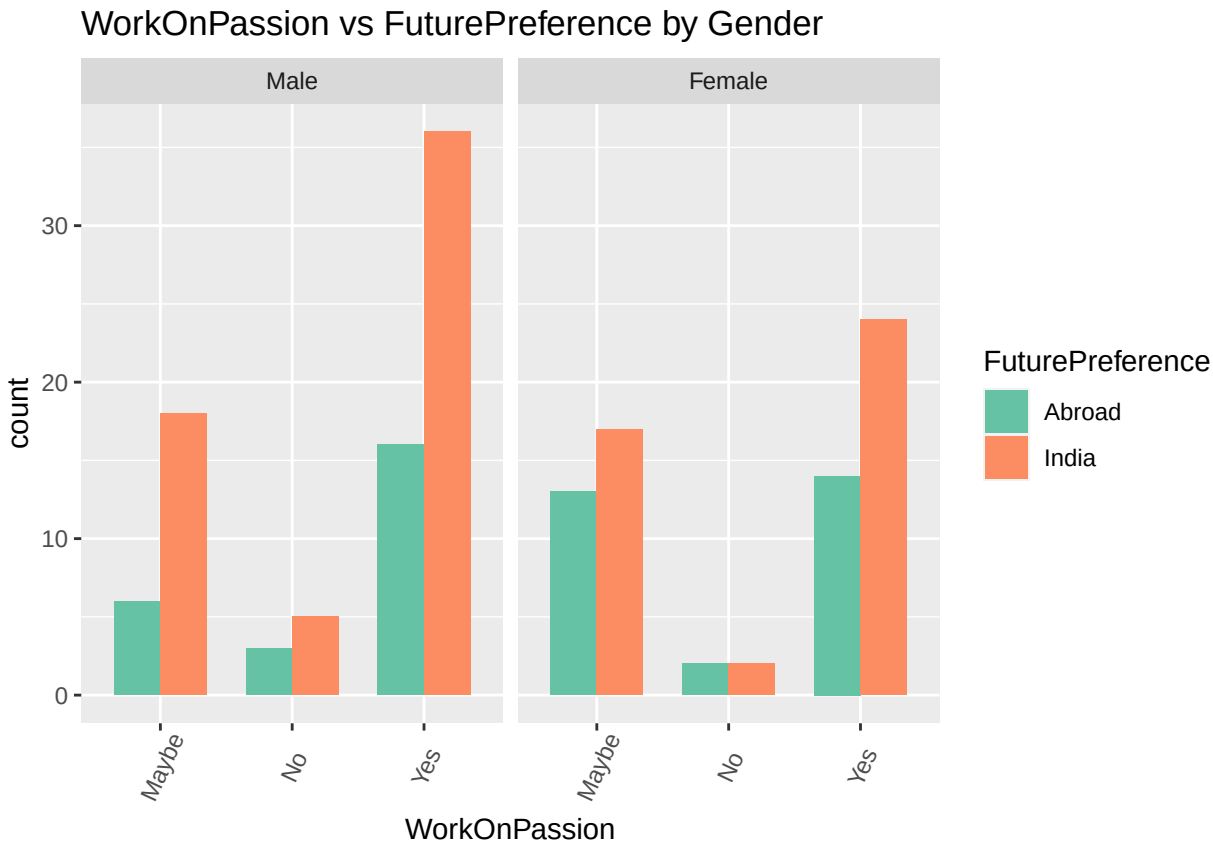



```
#=============================================================================
# INTERPRETATION :- The graph shows that there are distinct preferences for job
#types based on gender and educational stream. Men in Science and Commerce
#streams generally prefer higher-paying jobs, while women in Science, Commerce
#and as well as Arts streams prefer passion-related jobs. These preferences
#seem to reflect expectations and cultural norms regarding gender roles and
#career aspirations.
#=============================================================================

# Calculating mean age from age ranges
calculate_mean_age <- function(age_range) {
  ages <- strsplit(age_range, "-")[[1]]
  mean(as.numeric(ages))
}

# Clean the Age column by calculating mean ages from ranges
passion_clean <- passion %>%
  mutate(Age = sapply(Age, calculate_mean_age))

# Calculating mean age for each Occupation and Gender
mean_age <- passion_clean %>%
  filter(!is.na(Age)) %>%
  group_by(Occupation, Gender) %>%
  summarise(mean_age = mean(Age))
```

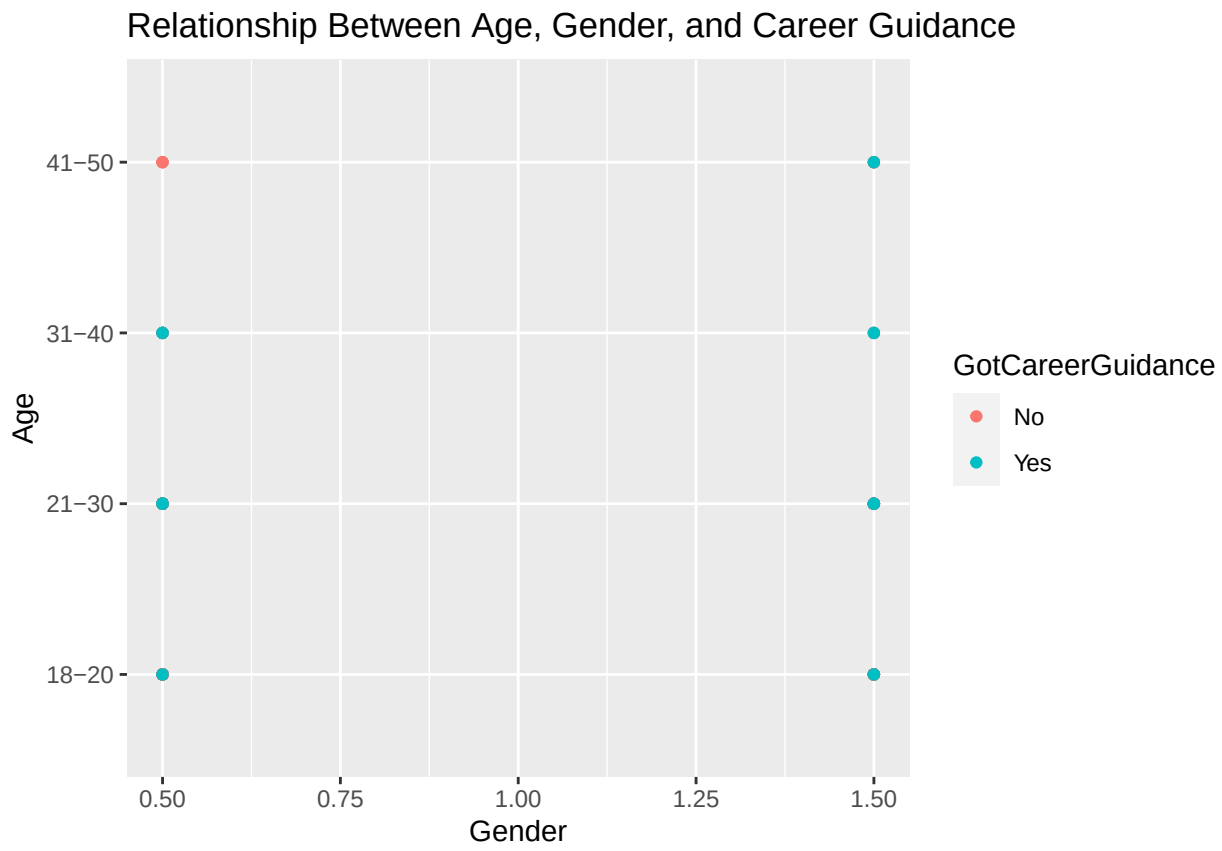
```
## `summarise()` has grouped output by 'Occupation'. You can override using the `.groups` argument.
```

```r
# Create line plot for Age Distribution by Occupation by Gender
ggplot(mean_age, aes(x = Occupation, y = mean_age, color = Gender,
                     group = Gender)) +
  geom_line() +
  labs(title = "Mean Age Distribution by Occupation by Gender",
       y = "Mean Age") +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6)) +
  scale_color_brewer(palette = "Set2")
```

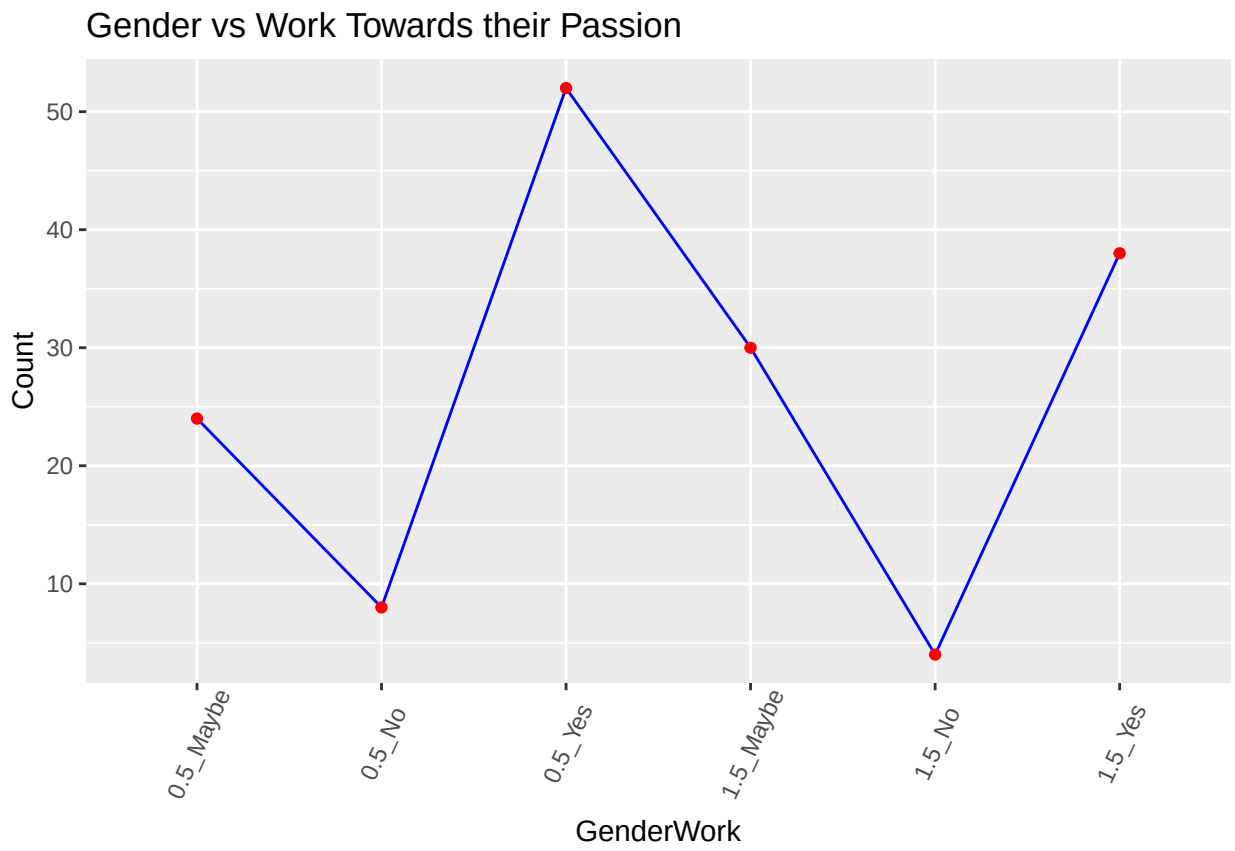

```r
#=============================================================================
# INTERPRETATION :- The graph shows that the average age of working
#professionals is higher than the average age of students. This is because
#working professionals have already completed their education and are
#established in their careers, while students are still in the process of
#learning and preparing for their future careers.
#=============================================================================

# Plotting WorkOnPassion vs FuturePreference by Gender
ggplot(passion, aes(x = WorkOnPassion, fill = FuturePreference)) +
  geom_bar(position = "dodge", width = 0.7) +
  labs(title = "WorkOnPassion vs FuturePreference by Gender") +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6)) +
  facet_wrap(~ Gender) +  # Adding facets for different genders
  scale_fill_brewer(palette = "Set2")
```




```
#=============================================================================
# INTERPRETATION :- More men than women find their current job or course to be
#their passion, while more women than men prefer to work or study abroad. This
#suggests that men are more likely to value passion in their work, while women
#are more likely to prioritize factors such as career opportunities and
#financial stability when making decisions about their future.
#=============================================================================

# Plotting Bubble chart of the relationship between age, gender, and career
#guidance
passion$Gender <- as.numeric(passion$Gender) - 0.5

ggplot(passion, aes(x = Gender, y = Age, color = GotCareerGuidance)) +
  geom_point() +
  labs(title = "Relationship Between Age, Gender, and Career Guidance")
```




```
#=============================================================================
# INTERPRETATION :- Older people are more likely to have received career
#guidance than younger people, and women are slightly more likely to have
#received career guidance than men. This suggests that career guidance is an
#important resource for people of all ages and genders, but it is especially
#beneficial for younger people and women.
#=============================================================================

#linegraph Gender vs Work Towards their Passion

TempWork = data.frame(matrix(nrow = 156, ncol = 2))
TempWork['GenderWork'] = paste0(passion$Gender, "_", passion$WorkOnPassion)

TempWork <- TempWork %>%
  group_by(GenderWork) %>%
  summarise(Count = n()) %>%
  ungroup()

g <- ggplot(TempWork, aes(x = GenderWork, y = Count, group = 1)) +
  geom_line(color = "blue", ) +
  geom_point(color = "red", ) +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6)) +
  labs(title = "Gender vs Work Towards their Passion")
g
```




```
#=============================================================================
# INTERPRETATION :- The graph suggests that there is no significant difference
#between the percentage of men and women who are working towards their passion
#because the percentage of men and women who are working towards their passion
#is very similar but more men than women find their current job or course to be
#their passion.
#=============================================================================

#=============================================================================
#=============================================================================
# OVERALL INTERPRETATION :- Based on the PASSION PURSUIT SURVEY DATASET,The
#frequency of people who are interested in pursuing a passion varies depending
#on their age, gender, and education stream. The percentage of people who are
#interested in pursuing a passion is highest among those aged 18-20, male, and
#from a science background. The educational stream choice varies with
#occupation, with working professionals more likely to choose streams related
#to their fields of study. Men typically opt for Science and Commerce streams,
#whereas women prefer Arts, Science, and Commerce streams. Occupation appears
#to have a more significant influence on educational preferences than gender.
#The gender disparity in career aspirations is not always consistent although
#there are some gender-based preferences. Men are more likely to choose STEM
#fields than women . In terms of job preferences, women prioritize higher-paying
#jobs while men prioritize passion-related jobs. Older people more likely to
#receive career guidance and are working towards their passion. Finally, older
#people tend to have a higher median age than younger people in this survey.
#=============================================================================
```

```
#==========================================================================
```