VDonde_A1

VaishnaviD

2024-02-11

Summary

The data tells us about how people in my class handle their drinks and food spendings along with their movie genre preference and their years of work experience.
Most people seem to stick to a moderate drinking routine, having around 1.5 drinks per week on average. But there are some who go a bit overboard, having up to 6 drinks in the same time frame.
When it comes to food, the spending habits are scattered all over the graph. Some don't spend anything at all, while others spend out up to $432 each week. On average people seem to spend about $37.50 on food every week.

The table provides insights into how people's experience levels influence their genre preferences.
Experience levels are denoted by letters:
L for 0-2 years of experience
M for 2-5 years of experience
H for more than 5 years of experience
Movie Genre are denoted by letters:
A = Action
H = Horror
C = Comedy
O = Other
Individuals with extensive experience (H) tend to favor action, comedy, and horror movies, indicating a preference for diverse and engaging genres. In contrast, those with less experience (L) show a variety of preferences across different genres, suggesting a more exploratory approach to movie-watching. Similarly, individuals with moderate experience (M) also display mixed genre preferences, indicating a transitional phase where tastes may still be evolving. Overall, the table highlights that while experience may shape genre preferences to some extent, individual tastes remain diverse and multifaceted.

Note

I have cleaned the data outside the R Programming in CSV Files so I have attached all the files with the code.Please download all the CSV files when you run the code in your system.

Analysis

1 - I observed that the data has both numerical and categorical variables and those are as mentioned below - Categorical variables - genre and experience Numerical variables - drinks and food

2 - The box plots illustrate variations in drinks and food consumption across different genres.

3 - I created a two-way table using the categorical variables genre and experience.It provides a summary of the frequency counts of combinations of genre and experience.This analyzes the distribution of experiences within each genre category.

To load necessary libraries.

```
library(ggplot2)
```

Data

```
# Read and combine all the data into a single dataframe
# Get list of csv files
files <- list.files(pattern = "\\.csv$")

# Initialize a dataframe
combined_data <- data.frame()

# Loop through each file and rbind to combined_data
for (file in files) {
  data <- read.csv(file)
  combined_data <- rbind(combined_data, data)
}

combined_data
```

```
##    genre experience drinks food
## 1      O          L      2  200
## 2      H          L      1   70
## 3      C          L      1  245
## 4      A          L      1  245
## 5      A          L      2   90
## 6      A          L      2   45
## 7      A          L      4   40
## 8      C          M      3   15
## 9      A          M      2   50
## 10     A          M      2   50
## 11     A          H      6  100
## 12     A          L      4   32
## 13     C          M      3   22
## 14     A          L      2   25
## 15     H          L      2   15
## 16     C          M      2   40
## 17     C          L      2   20
## 18     C          H      6   50
## 19     O          L      1   30
## 20     A          L      0   20
## 21     A          H      1   35
## 22     A          H      3    0
## 23     H          M      2   50
## 24     A          H      1  400
## 25     A          H      1  432
## 26     C          L      0   30
## 27     C          L      2   20
## 28     C          H      6   50
## 29     O          L      1   30
## 30     A          L      0   20
## 31     A          H      3   35
## 32     A          L      0   20
## 33     A          H      2  300
## 34     C          L      0   19
## 35     C          M      0   20
## 36     A          M      1   70
## 37     C          L      0  250
## 38     A          L      1  150
## 39     A          L      1  120
## 40     H          L      0  100
## 41     O          L      0  100
## 42     O          L      2   15
## 43     C          M      0    0
## 44     C          L      2   10
## 45     O          M      1   20
## 46     H          H      1   10
```

Perform basic descriptive statistics on the variables.

```
summary(combined_data)
```

```
##     genre            experience           drinks          food       
##  Length:46          Length:46          Min.   :0.000   Min.   :  0.00  
##  Class :character   Class :character   1st Qu.:1.000   1st Qu.: 20.00  
##  Mode  :character   Mode  :character   Median :1.500   Median : 37.50  
##                                        Mean   :1.717   Mean   : 80.65  
##                                        3rd Qu.:2.000   3rd Qu.: 97.50  
##                                        Max.   :6.000   Max.   :432.00
```

Explaining the data type.

genre - Categorical data
experience - Categorical data
drinks - Quantitative data
food - Quantitative data

```
str(combined_data)
```

```
## 'data.frame':    46 obs. of  4 variables:
##  $ genre     : chr  "O" "H" "C" "A" ...
##  $ experience: chr  "L" "L" "L" "L" ...
##  $ drinks    : int  2 1 1 1 2 2 4 3 2 2 ...
##  $ food      : int  200 70 245 245 90 45 40 15 50 50 ...
```

For the numerical variables,to determine the 5-number summary.

```
fivenum(combined_data$drinks)
```

```
## [1] 0.0 1.0 1.5 2.0 6.0
```

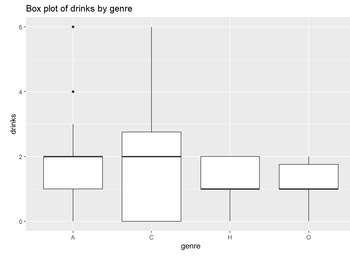
```
fivenum(combined_data$food)
```

```
## [1]   0.0  20.0  37.5 100.0 432.0
```

To plot box plots for each of the continuous variables by Genre.

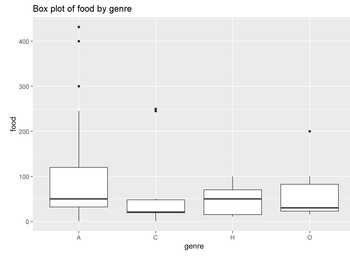
a)Box plot of drinks by genre.

```
ggplot(combined_data, aes(x = genre, y = drinks)) +
  geom_boxplot() +
  labs(title = "Box plot of drinks by genre")
```



b)Box plot of food by genre.

```
ggplot(combined_data, aes(x = genre, y = food)) +
  geom_boxplot() +
  labs(title = "Box plot of food by genre")
```



To plot a two-way table using the two categorical variables.

```
table(combined_data$genre, combined_data$experience)
```

```
##
##     H  L  M
##   A  7 11  3
##   C  2  7  5
##   H  1  3  1
##   O  0  5  1
```