Kiran Kasu

ALY6000

Prof. Ji-Young Yun

Final Report

Contents

Introduction

This final report is about creating our own dataset. The class has created individual surveys to collect the data from the remaining class members. The survey responses are considered, and a data analysis plan has been created. The data visualization and interpretations have been included in this report for better analysis.

The report shows the introduction, analysis, conclusion, and the references used to complete this report with the R code and script used to do the visualizations.

Analysis

Levels of Measurement : To perform statistical analysis of data, it is important to first understand variables and what should be measured using these variables. There are different levels of measurement in statistics and data measured using them can be broadly classified into qualitative and quantitative data.

Measurement Scales used:

- Nominal Scale
- Ordinal Scale
- Interval Scale
- Ratio Scale

Nominal Scale

- It is also called the categorical variable scale; it is defined as a scale used for labeling variables into distinct classifications and doesn't involve a quantitative value or order. This scale is the simplest of the four variable measurement scales.
- I have collected the gender data as the nominal scale data for the survey created.
- Nominal scale is a naming scale, where variables are simply "named" or labeled, with no specific order.
- There are two primary ways in which nominal scale data can be collected:

1. By asking an open-ended question, the answers of which can be coded to a respective number of label decided by the researcher.

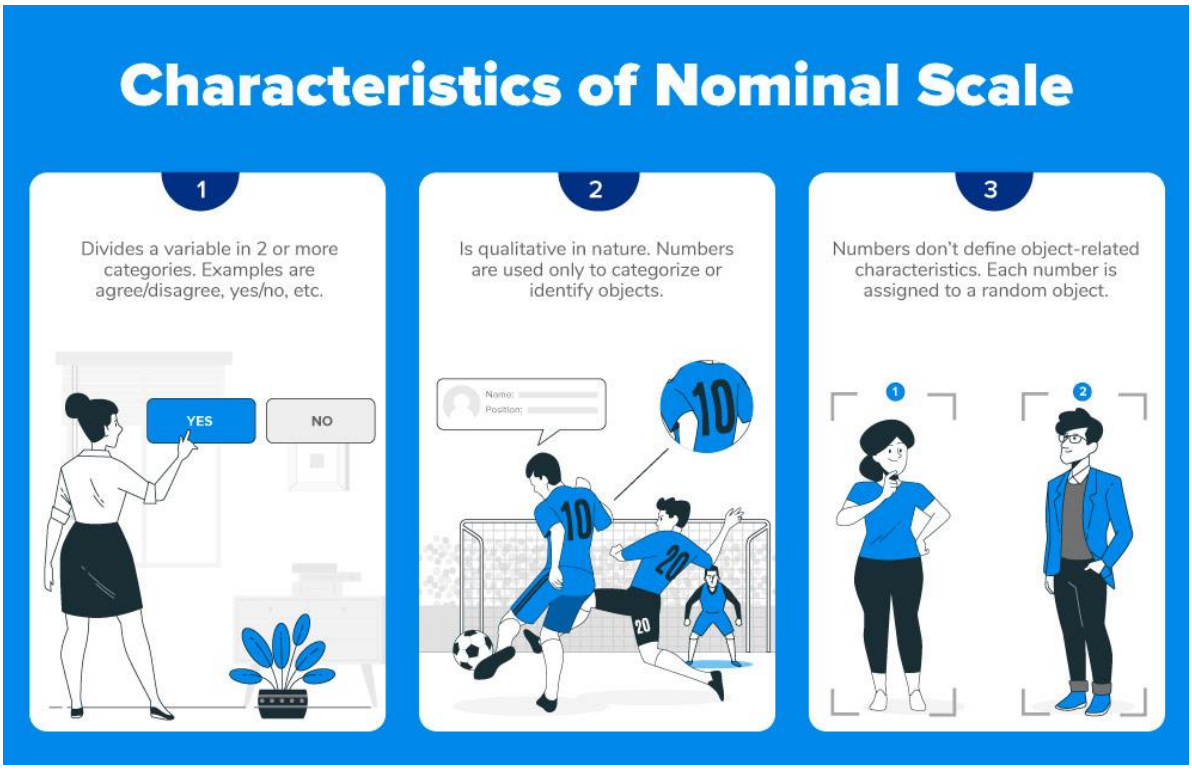
2. The other alternative to collect nominal data is to include a multiple-choice question in which the answers will be labeled.

Nominal Scale Examples

1. Gender

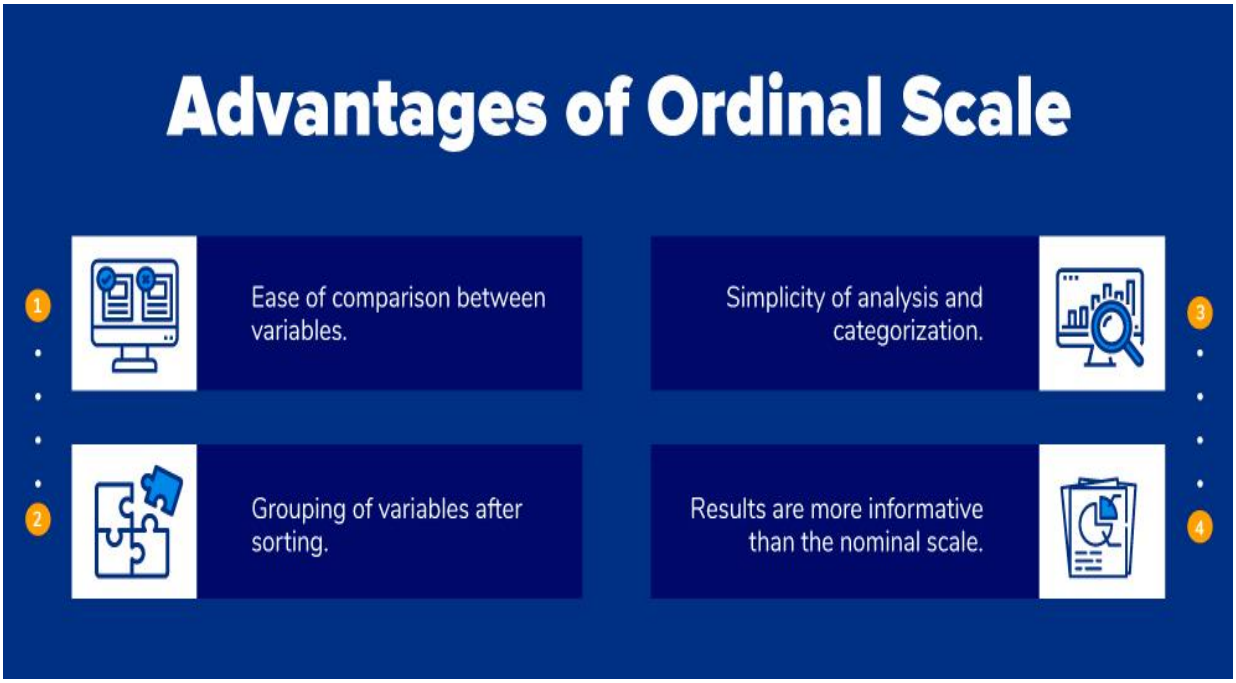
2. Political preferences

3. Place of residence



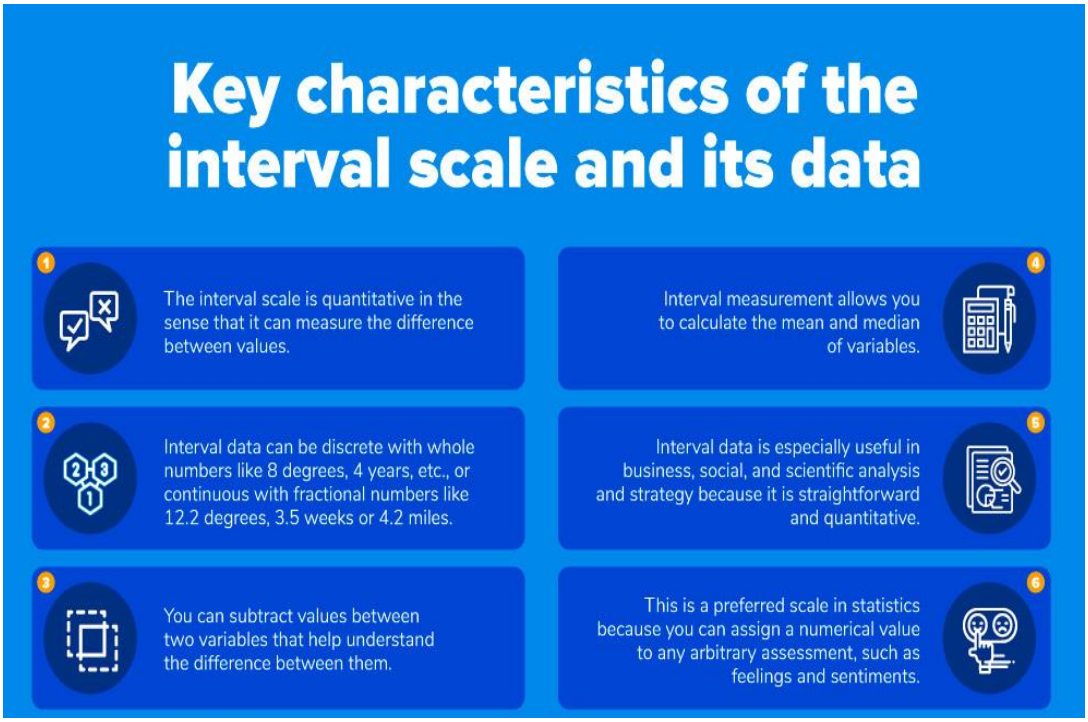
Ordinal Scale

- It is defined as a variable measurement scale used to simply depict the order of variables and not the difference between each of the variables.
- Ordinal Scale maintains descriptional qualities along with an intrinsic order but is void of an origin of scale and thus, the distance between variables can't be calculated.
- Ordinal Scale data can be presented in tabular or graphical formats for a researcher to conduct a convenient analysis of collected data.
- Status at workplace, tournament team rankings, order of product quality, and order of agreement or satisfaction are some of the most common examples of the ordinal Scale.
- Ordinal Scale examples:
    1. Grades
    2. Satisfaction
    3. Happiness



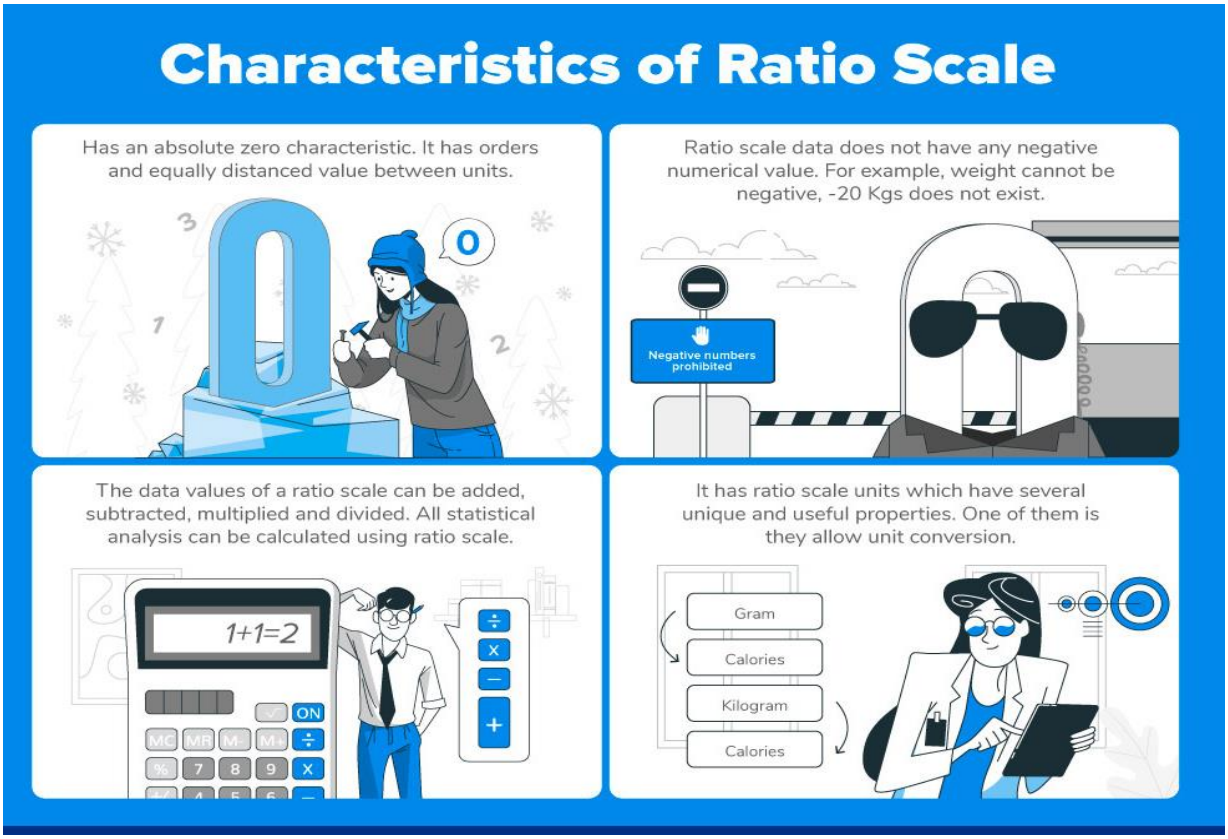
Interval Scale

- It is defined as a numerical scale where the order of the variables is known as well as the difference between these variables.
- I have collected the data of how far is the nearby restaurant from your house as a interval scale for the survey created .
- These scales are effective as they open doors for the statistical analysis of provided data.
- All the techniques applicable to nominal and ordinal data analysis are applicable to Interval Data as well.
- Apart from those techniques, there are a few analysis methods such as descriptive statistics, correlation regression analysis which is extensively for analyzing interval data.
- Interval Scale examples:
    1. What is your family income ?
    2. What is the temperature in your city ?



Ratio Scale

- It is defined as a variable measurement scale that not only produces the order of variables but also makes the difference between variables known along with information on the value of true zero.
- I have collected the data of amount spent on food per day as a ratio scale for the survey created
- These scales are effective as they open doors for the statistical analysis of provided data.
- At a fundamental level, Ratio scale data is quantitative in nature due to which all quantitative analysis techniques such as SWOT, TURF, Cross-tabulation, Conjoint, etc. can be used to calculate ratio data.
- Ratio Scale examples:
    1. What is your current height ?
    2. What is your weight ?



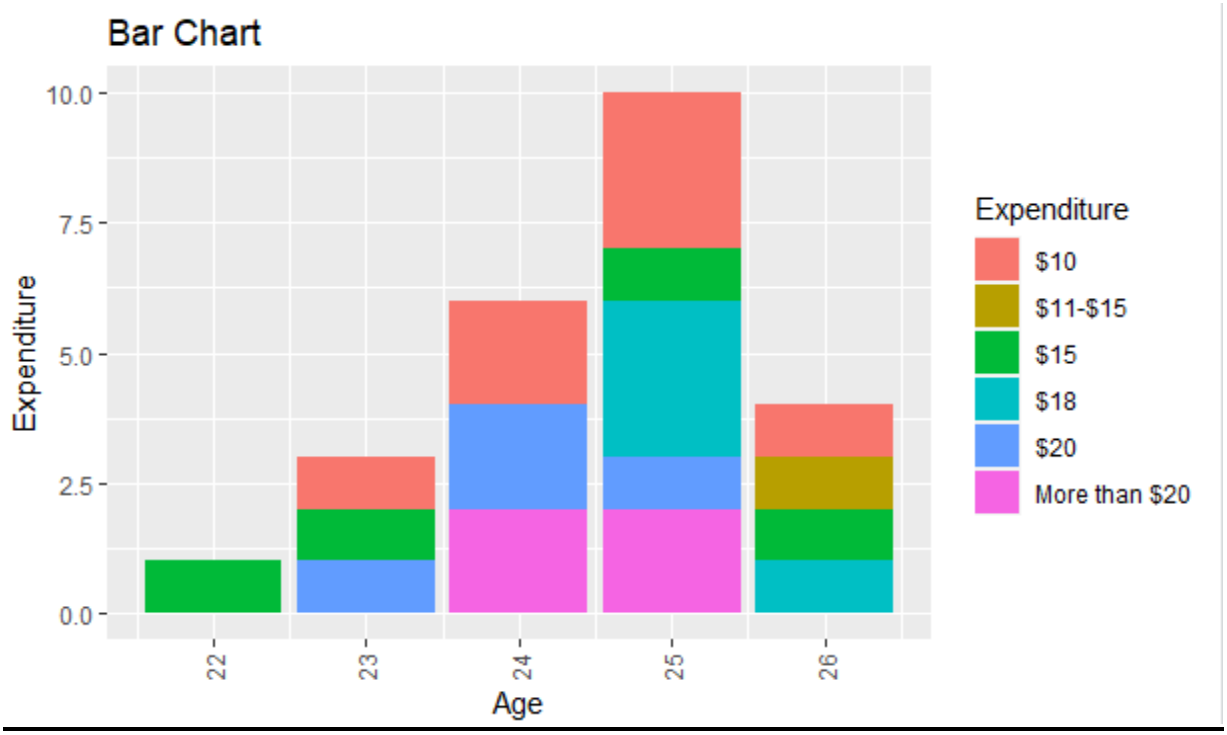
The survey conducted consisted of few attributes with different data types hence made a few changes to responses to complete the analysis of the report.

1. The first analysis is a stacked bar graph with gender and age as the attributes in X and Y axis respectively. Here in this analysis we have found out from the responses that are provided were represented in the Stacked Bar Chart with the age and the expenditure of the class members spent on food per day.

   Command:

   ggplot(data, aes(Age, fill = Expenditure)) + geom_bar()+ labs(title = "Bar Chart", x = "Age", y = "Expenditure")+theme(axis.text.x = element_text(angle = 90, vjust = 0.5))

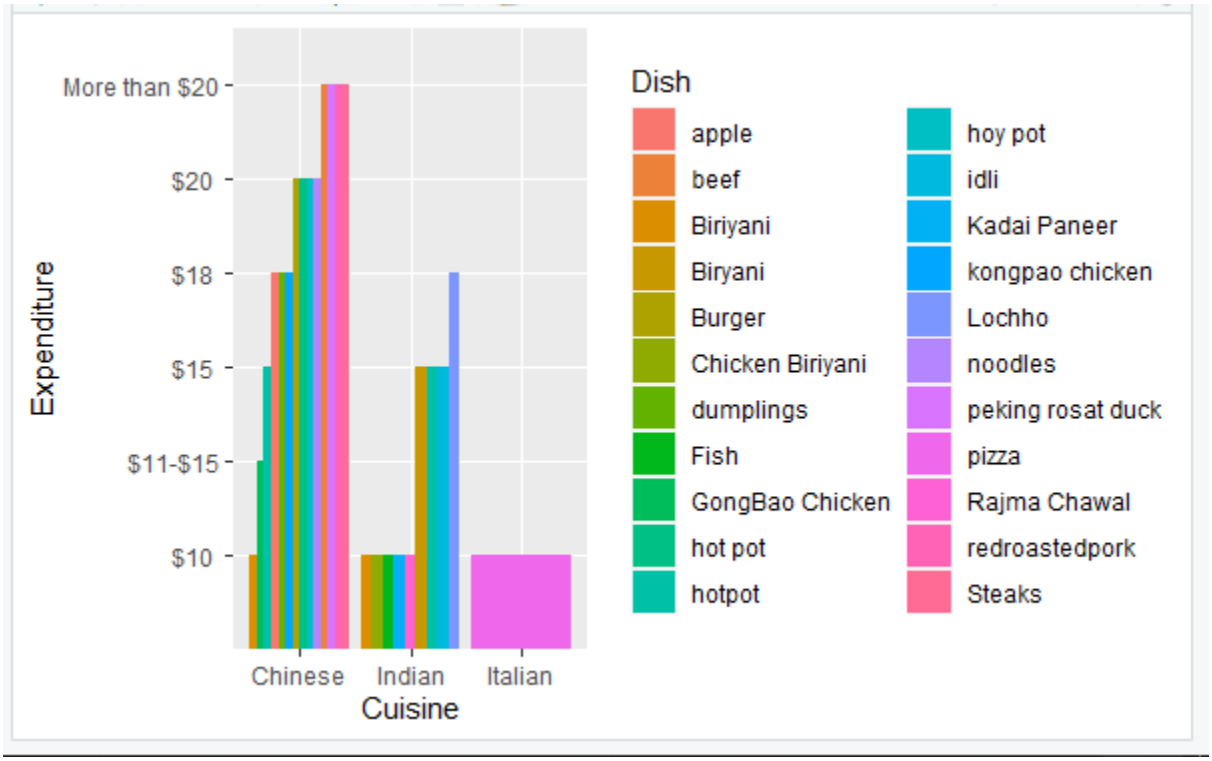
Output:



2. The bar plot representation of the cuisine and the Expenditure of the class members with in the survey conducted and the responses received are represented in the form of a bar graph. We can see the cuisine selected along with the amount spent per day by the students.

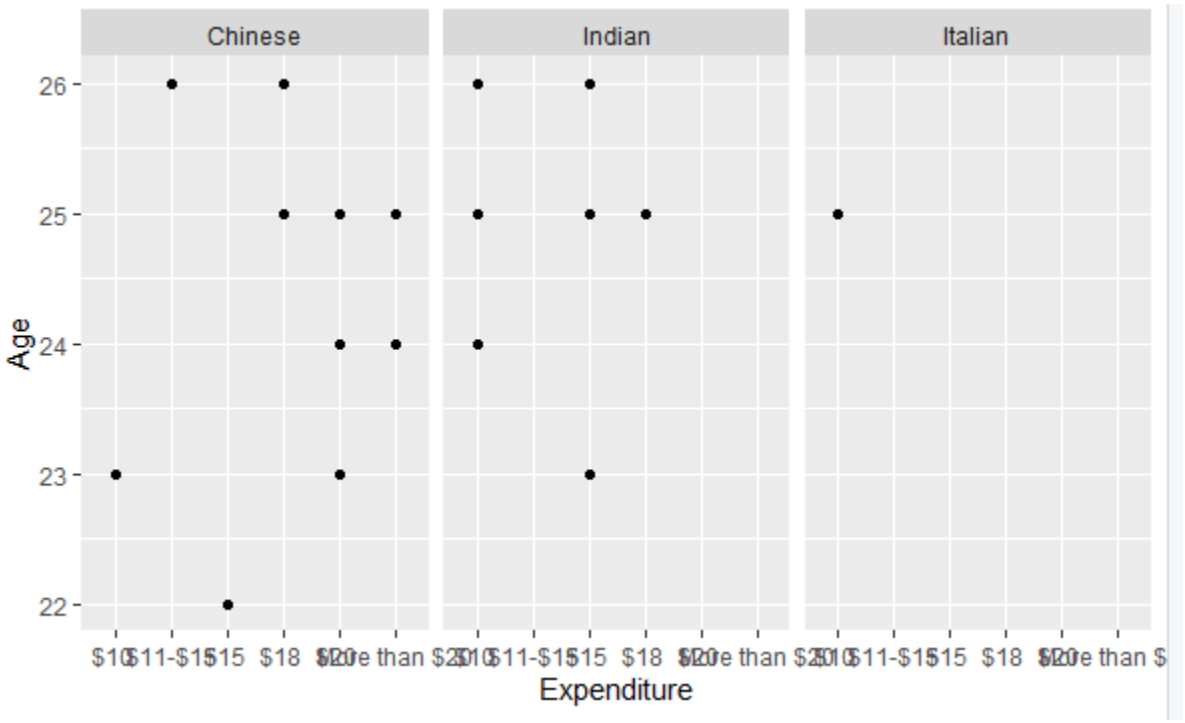
Command: ggplot(data, aes(Cuisine,Expenditure, fill = Dish)) + geom_col(position = "dodge")

Output:



3. Scatter Plot representation with the expenditure and the cuisine is to be seen in the below graph. With in the 25 responses received we can see there are many students who like Chinese and Indian cuisine more.
   Command: ggplot(data = df, aes(x = Expenditure, y = Age)) + geom_point() + facet_wrap(~Cuisine)

Output:

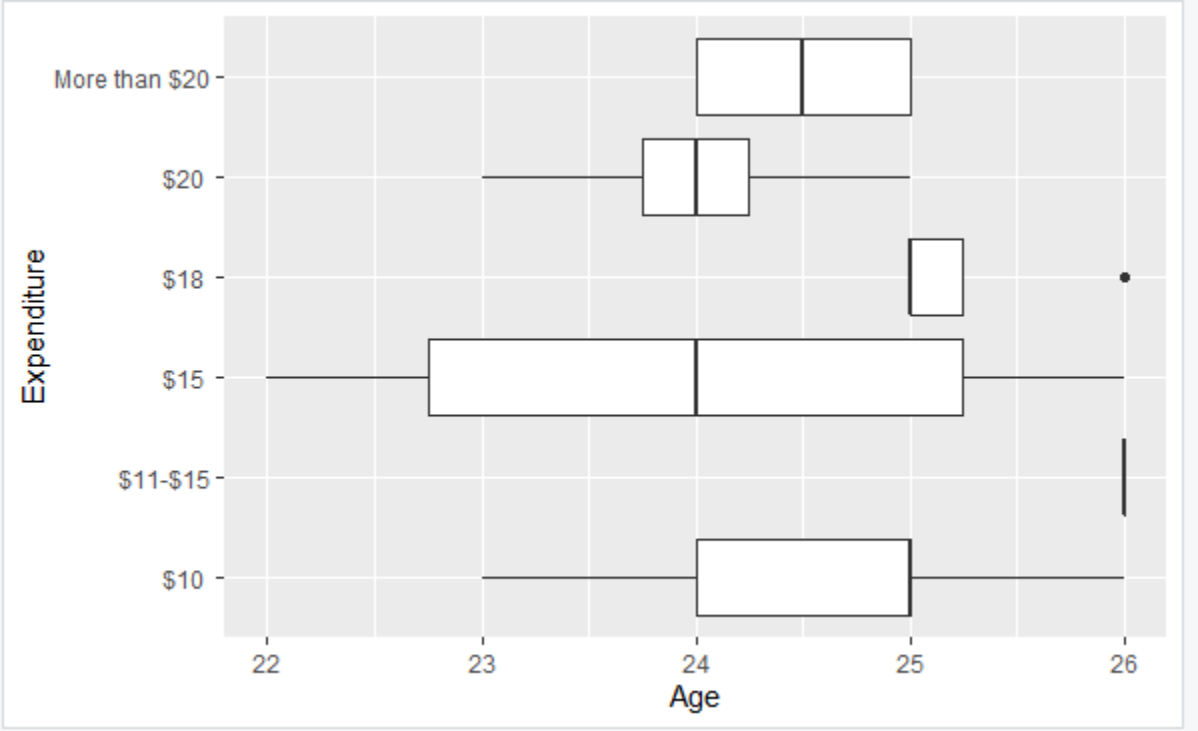


4. Bar Plot for the expenditure and age can been seen in the below graph accordingly to the age of the students to the expenditure they spend as a lot of students spend an average amount of $15 dollars per day on food.

Command:

ggplot(data = df, aes(x = Expenditure, y = Age)) + geom_boxplot() + coord_flip()

Output:



5. Correlation analysis for the age and expenditure has been represented here

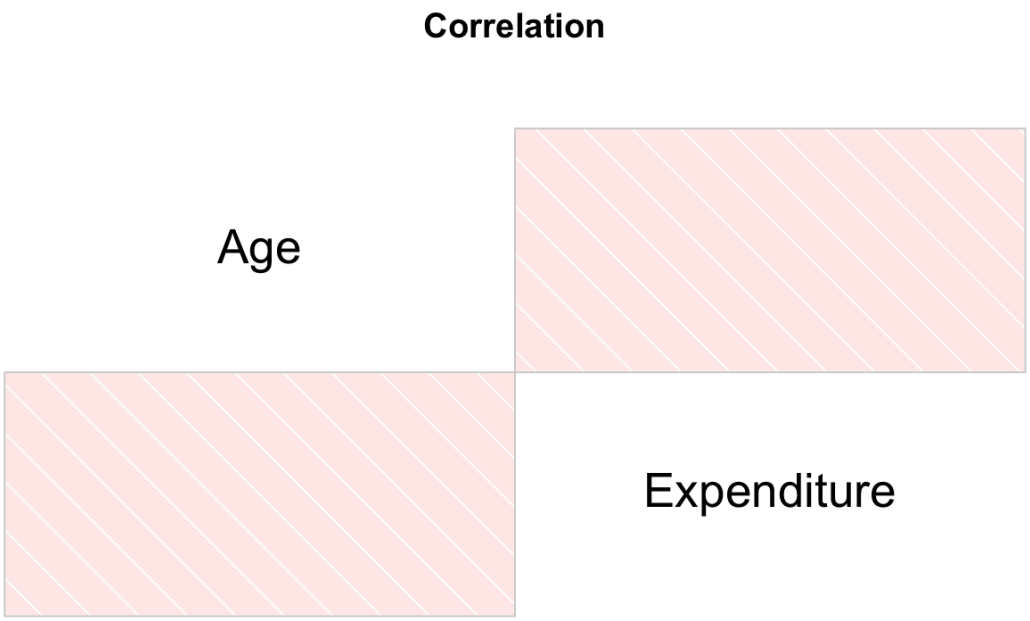
Command:

```
install.packages("corrgram", repos ="https://cran.rstudio.com/bin/macosx/big-sur-arm64/contrib/4.1/corrgram_1.14.tgz")

library(corrgram)

corrgram(datasetn, order=NULL,

        panel=panel.shade, text.panel=panel.txt,

        main="Correlation")

#coorelation
```

Expenditure = c(15,10,10,15,15,24,18,18,18,22,18, 25, 10, 15, 20,10,20,23,10,10,20,14,20,10)

datasetn$Expenditure = Expenditure

Output:



Conclusion

- Knowing the measurement scale for your variables can help prevent mistakes like taking the average of a group of zip (postal) codes, or taking the ratio of two pH values.

- When it comes to creating an analysis plan for your project, Begin with simple methods to summarize and visualize your data.

- Quantitative research deals with numbers and statistics, while qualitative research deals with words and meanings.

- The survey conducted was to know the food preferences of the class and the responses received are to be analyzed and visualized.

- Following the measurement scales used in the survey to collect data and the computation factors.

References

- *Nominal, ordinal, interval, ratio scales with examples*. (2021, April 27). QuestionPro. https://www.questionpro.com/blog/nominal-ordinal-interval-ratio/

- *Types of data measurement scales: Nominal, ordinal, interval, and ratio*. (2020, October 5). My Market Research Methods. https://www.mymarketresearchmethods.com/types-of-data-nominal-ordinal-interval-ratio/

- *What is the difference between ordinal, interval and ratio variables? Why should I care?* (n.d.). GraphPad. https://www.graphpad.com/support/faq/what-is-the-difference-between-ordinal-interval-and-ratio-variables-why-should-i-care/

Appendix:

R script:

```
install.packages("FSA")
install.packages("FSAdata")
install.packages("magrittr")
install.packages("dplyr")
install.packages("plotrix")
install.packages("ggplot2")
install.packages('plyr')
install.packages("tidyverse")
install.packages('data.table')
install.packages("plotly")

library("plyr")
library("data.table")
library("dplyr")
library("FSA")
library("FSAdata")
library("magrittr")
library("dplyr")
library("plotrix")
library("ggplot2")
library("plyr")
library("tidyverse")
library(plotly)
library(dplyr)
library("MASS")
#data set loading
data = read.csv("Survey.csv")
summary(data)

#plot 1 graph
library("ggplot2")

#plot 2 graph

ggplot(data, aes(Age, fill = Expenditure)) + geom_bar()+
  labs(title = "Bar Chart", x = "Age", y = "Expenditure")+theme(axis.text.x =
element_text(angle = 90, vjust = 0.5))
```

```
  #plot
  ggplot(data, aes(Cuisine,Expenditure, fill = Dish)) + geom_col(position = "dodge")


  #scatterplot
  ggplot(data = df, aes(x = Expenditure, y = Age)) + geom_point() +
facet_wrap(~Cuisine)


  #barplot
  ggplot(data = df, aes(x = Expenditure, y = Age)) + geom_boxplot() +   coord_flip()

  data = read.csv("Survey.csv")
  summary(data)

  install.packages("corrgram", repos ="https://cran.rstudio.com/bin/macosx/big-sur-
arm64/contrib/4.1/corrgram_1.14.tgz")
  library(corrgram)
  corrgram(datasetn, order=NULL,
         panel=panel.shade, text.panel=panel.txt,
         main="Correlation")
  #coorelation
  Expenditure = c(15,10,10,15,15,24,18,18,18,22,18, 25, 10, 15,
20,10,20,23,10,10,20,14,20,10)
  datasetn$Expenditure = Expenditure
```