

step 1

```
survey <- readxl::read_xlsx('05. 2024STB_survey_80.xlsx')
```

```
colnames(survey) <- c(
  "Gender",
  "Age",
  "Grade",
  "Nationality",
  "Residential_Area",
  "Internet_Usage_Last_Year",
  "Health_Satisfaction",
  "Income_Satisfaction",
  "Housing_Satisfaction",
  "Family_Relationship_Satisfaction",
  "Social_Relationship_Satisfaction",
  "Leisure_Satisfaction",
  "Overall_Life_Satisfaction",
  "Regular_Donation_Volunteer",
  "Donation_Amount_2023",
  "Volunteer_Activities_2022",
  "Parent_Separation_Status",
  "Parent_Visit_Frequency_2023",
  "Parent_Call_Frequency_2023",
  "Lifetime_Smoking_Amount",
  "First_Smoking_Age",
  "Total_Smoking_Years",
  "Current_Smoking_Status",
  "Daily_Smoking_Amount",
  "Quit_Smoking_Attempts_1Year",
  "Future_Quit_Smoking_Plans",
  "Secondhand_Smoke_Exposure_Hours",
  "Secondhand_Smoke_Exposure_Hours_Avg",
  "Alcohol_Frequency",
  "Drinks_Per_Session",
  "Heavy_Drinking_Frequency",
  "Inability_to_Stop_Drinking",
  "Uncompleted_Tasks_Due_to_Alcohol",
  "Morning_After_Drinking_Frequency",
  "Alcohol_Regret_Frequency",
  "Blackout_Frequency",
  "Injury_Due_to_Alcohol",
  "Concerns_from_Others_about_Alcohol"
)
```

Convert column

```
# Convert 'Gender' column: "남자 男人" -> 1 (Male), "여자 女子" -> 2 (Female)
survey$Gender <- ifelse(survey$Gender == "남자 男人", 1,
                        ifelse(survey$Gender == "여자 女子", 2, NA))
```

step 2

```
> # Frequency distribution for Gender
> gender_freq <- table(survey$Gender)
> gender_freq
```

```
 1  2
33 47
```

step 3

```
> # Relative frequency distribution for Gender
> gender_relative_freq <- prop.table(table(survey$Gender))
> gender_relative_freq
```

```
      1      2
0.4125 0.5875
```

step 4

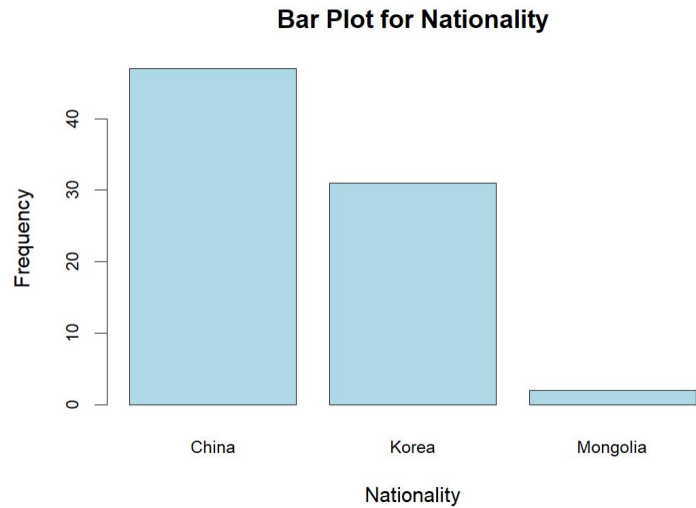
```
> survey$Grade <- ifelse(survey$Grade == "2학년 / 2年级", 'Grade 2',
+                         ifelse(survey$Grade == "3학년 / 3年级", 'Grade 3',
+                               ifelse(survey$Grade == "4학년 / 4年级", 'Grade 4',
+                                     ifelse(survey$Grade == "5학년 / 5年级", 'Grade 5', NA))))
> gender_grade_crosstab <- table(survey$Gender, survey$Grade)
> gender_grade_crosstab
```

```
      Grade 2 Grade 3 Grade 4 Grade 5
1         14      7      12      0
2          9     14     23      1
```

step 5

```
survey$Nationality <- ifelse(survey$Nationality == "한국 / 韓國", "Korea",
                             ifelse(survey$Nationality == "중국 / 中國", "China",
                                     ifelse(survey$Nationality == "몽골", "Mongolia", NA)))

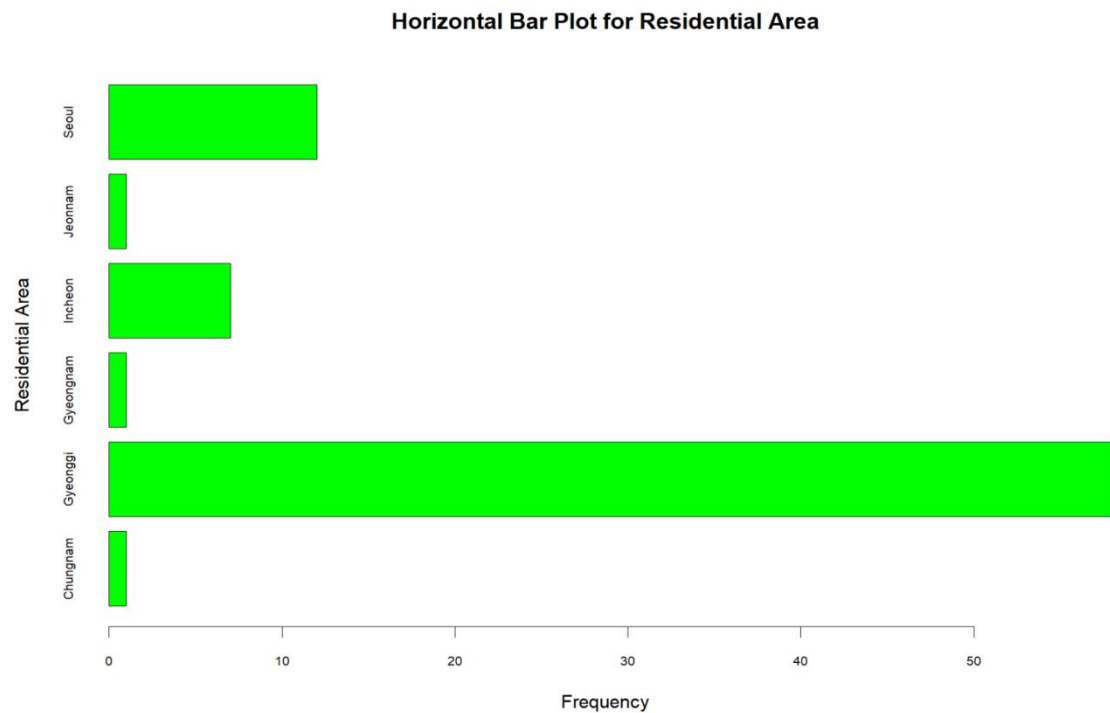
barplot(table(survey$Nationality),
        main = "Bar Plot for Nationality",
        xlab = "Nationality",
        ylab = "Frequency",
        col = "lightblue")
```



step 6

```
survey$Residential_Area <- ifelse(survey$Residential_Area == "경기 畿技", "Gyeonggi",
                                ifelse(survey$Residential_Area == "인천 仁川", "Incheon",
                                        ifelse(survey$Residential_Area == "전남 全南", "Jeonnam",
                                              ifelse(survey$Residential_Area == "서울 首尔", "Seoul",
                                                    ifelse(survey$Residential_Area == "충남 忠南", "Chungnam",
                                                          ifelse(survey$Residential_Area == "경남 庆南", "Gyeongnam", NA))))))

par(cex.axis = 0.7)
barplot(table(survey$Residential_Area),
        main = "Horizontal Bar Plot for Residential Area",
        xlab = "Frequency",
        ylab = "Residential Area",
        col = "#FF69B4",
        horiz = T)
```



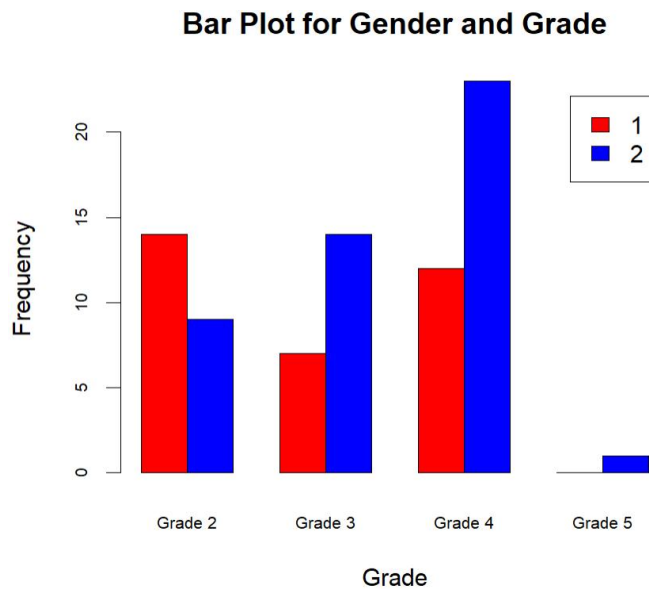
step 7

"남자 男人" -> 1 (Male), "여자 女子" -> 2 (Female)

```

barplot(table(survey$Gender, survey$Grade),
        beside = TRUE,
        legend = rownames(table(survey$Gender, survey$Grade)),
        main = "Bar Plot for Gender and Grade",
        xlab = "Grade",
        ylab = "Frequency",
        col = c("red", "blue"))

```

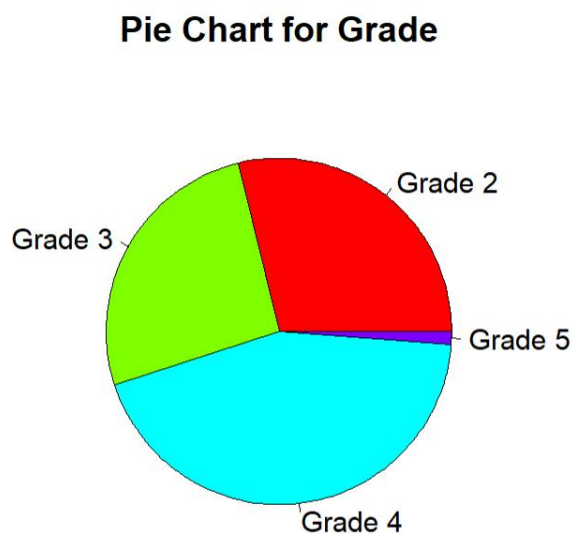


step 8

```

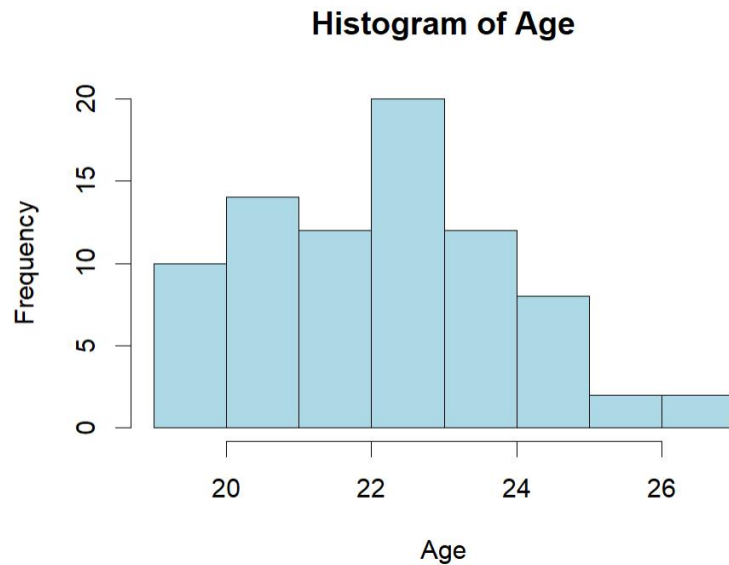
grade_freq <- table(survey$Grade)
pie(grade_freq,
    main = "Pie chart for grade",
    col = rainbow(length(grade_freq)),
    labels = names(grade_freq))

```

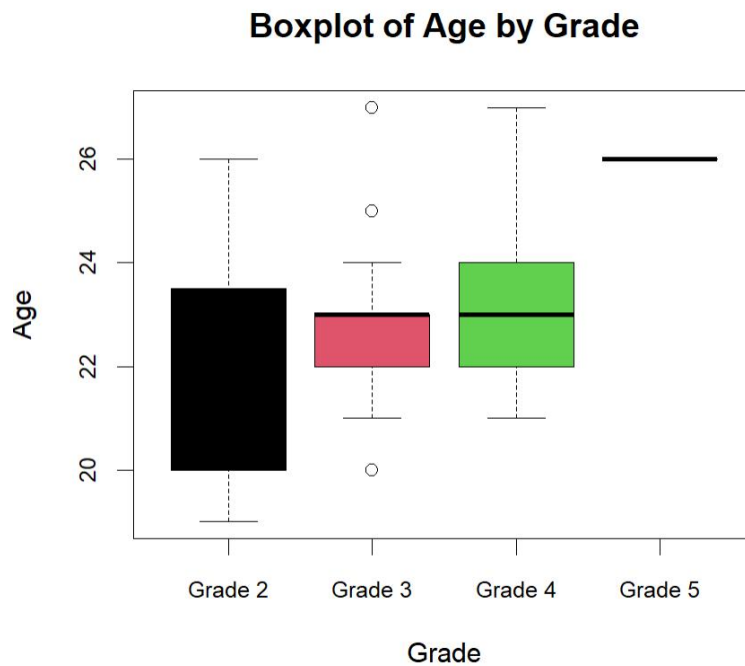


step 9

```
hist(as.numeric(survey$Age),
     main = "Histogram of Age",
     xlab = "Age",
     ylab = "Frequency",
     col = "lightblue")
```



```
# step 10
par(cex.axis = 0.8)
boxplot(as.numeric(survey$Age) ~ survey$Grade,
        main = "Boxplot of Age by Grade",
        xlab = "Grade",
        ylab = "Age",
        col = 1:4)
```



2학년 학생의 연령 분포는 19 세부터 26 세까지 더 넓습니다. 평균 연령은 21 세로 대부분의 학생이 이 연령에 속함을 나타냅니다. 사분위간 범위 (IQR)가 커서 이 학년 수준에서 연령의 분산이 더 높습니다.

3학년 학생의 연령은 22~23 세에 집중되어 있으며, 사분위수를 보면 학생의 연령이 주로 이 범위에 집중되어 있음을 알 수 있습니다. IQR은 상대적으로 작아서 이 학년에 연령 분포가 집중되어 있음을 나타냅니다.

4학년의 연령은 3학년의 연령과 유사하며 연령은 22 세에서 24 세 사이에 밀집되어 있으며 사분위수 범위가 약간 더 넓다는 것은 연령 분포가 3학년보다 약간 더 분산되어 있음을 나타냅니다.

5학년의 모든 학생은 26 세이며 연령 변화는 없습니다. 이는 모든 학생의 나이가 동일하고 모든 데이터 포인트가 동일하기 때문에 상자 그림이 IQR 없이 평평한 선으로 나타남을 의미합니다.

step 11

```
survey$Grade_numeric <- ifelse(survey$Grade == "2학년 / 2年级", 2,
                               ifelse(survey$Grade == "3학년 / 3年级", 3,
                                       ifelse(survey$Grade == "4학년 / 4年级", 4,
                                             ifelse(survey$Grade == "5학년 / 5年级", 5, NA))))

# scatter plot
plot(as.numeric(survey$Age), survey$Grade_numeric,
     main = "Scatter Plot of Age vs Grade",
     xlab = "Age",
     ylab = "Grade (numeric)",
     pch = 20)
```

Scatter Plot of Age vs Grade

