# step 1

survey <- readxl::read\_xlsx('05. 2024STB\_survey\_80.xlsx')

**# Redefine column names**

· **Gender** → Gender

· **Age** → Age

· **Grade** → Grade

· **Nationality** → Nationality

· **residential area** → Residential\_Area

· **6. 귀하는 인터넷을 작년에 인터넷을 사용한 적이 있습니까? / 6. 您去年用过互联网吗？** → Internet\_Usage\_Last\_Year

· **7. 본인의 건강상태에 대한 만족도는 어느 정도입니까? / 7. 您对你的健康状况的满意度是多少?** → Health\_Satisfaction

· **8. '가족의 수입'에 대한 귀하의 만족도는 어느 정도입니까? / 8. 您对"家庭收入"的满意度是多少?** → Income\_Satisfaction

· **9. 본인의 '주거 환경'에 대한 귀하의 만족도는 어느 정도입니까? / 9. 您对自己的"居住环境"满意吗?** → Housing\_Satisfaction

· **10. 본인의 '가족 관계'에 대한 귀하의 만족도는 어느 정도입니까? / 10. 您对本人的"家庭关系"满意程度如何？** → Family\_Relationship\_Satisfaction

· **11. 본인의 '사회적 친분관계'에 대한 귀하의 만족도는 어느 정도입니까? / 11. 您对本人的"社会交情"满意程度如何？** → Social\_Relationship\_Satisfaction

· **12. 본인의 '여가생활'에 대하여 귀하의 만족도는 어느 정도입니까? / 12. 您对本人的"休闲生活"满意吗？** → Leisure\_Satisfaction

· **13. 그럼, 위의 사항을 모두 고려할 때, 귀하는 전반적으로 자신의 생활에 얼마나 만족하고 계십니까? /13. 那么，考虑到以上情况，您对自己的生活总体上还满意吗？** → Overall\_Life\_Satisfaction

· **14. 귀하는 정기적으로 기부를 하거나, 자원봉사활동을 하고 있습니까? / 14. 你有定期捐款或者参加志愿活动吗？** → Regular\_Donation\_Volunteer

· **14-1. (문14)의 '그렇다' 응답자만) 귀하는 2023년 1년간 얼마 정도(연간 총액, 단위: 만원) 기부하셨습니까? 14-1. (文14中"是"的回答者)您在2023年1年间捐赠了多少(年总额,单位:万元)?** → Donation\_Amount\_2023

· **(문14)의 '그렇다' 응답자만) 귀하는 자원봉사활동을 2022년 1년간 몇 회 정도 하셨습니까? / (文14中"是"的受访者)您在2022年一年内进行了多少次志愿服务活动?** → Volunteer\_Activities\_2022

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"

)

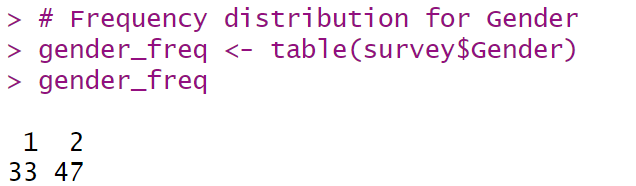
**# Convert column**

# Convert 'Gender' column: "남자 男人" -> 1 (Male), "여자 女子" -> 2 (Female)

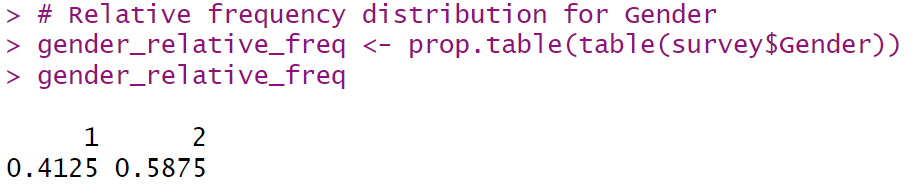
survey$Gender <- ifelse(survey$Gender == "남자 男人", 1,

ifelse(survey$Gender == "여자 女子", 2, NA))

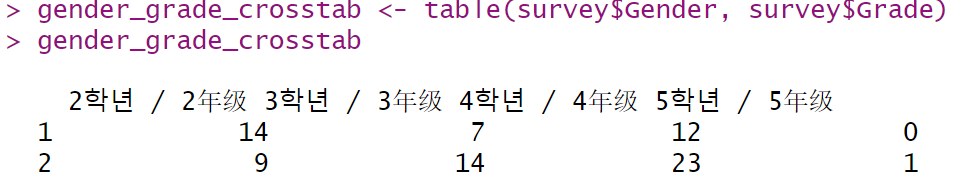
# step 2



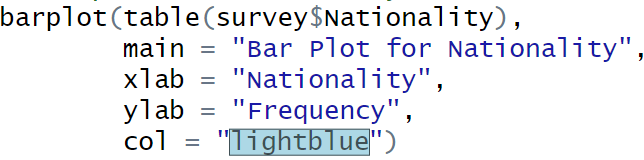
# step 3

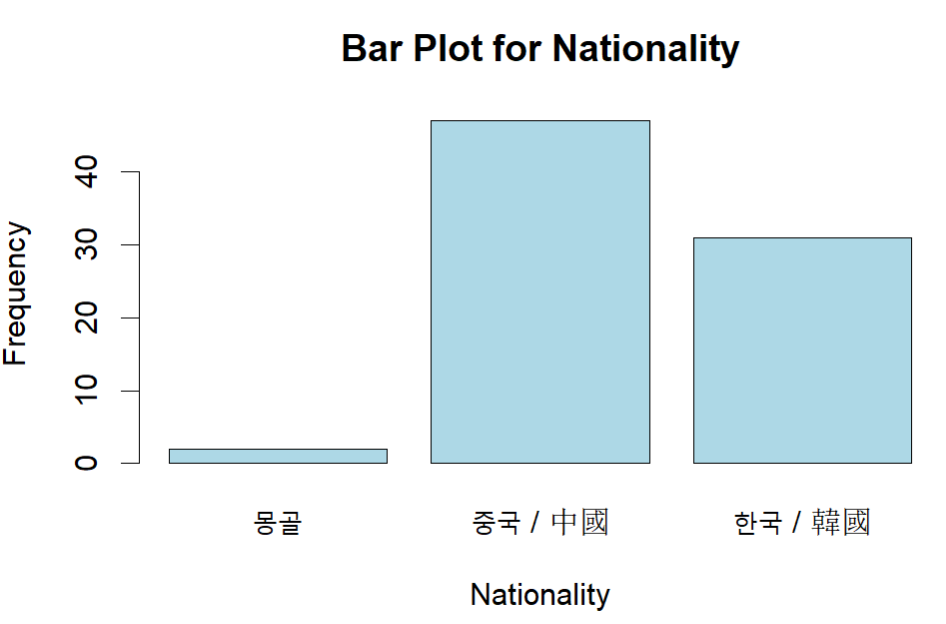


# step 4

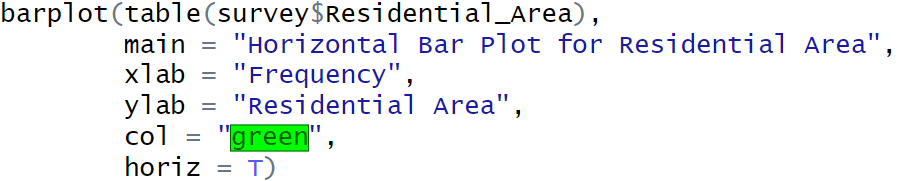


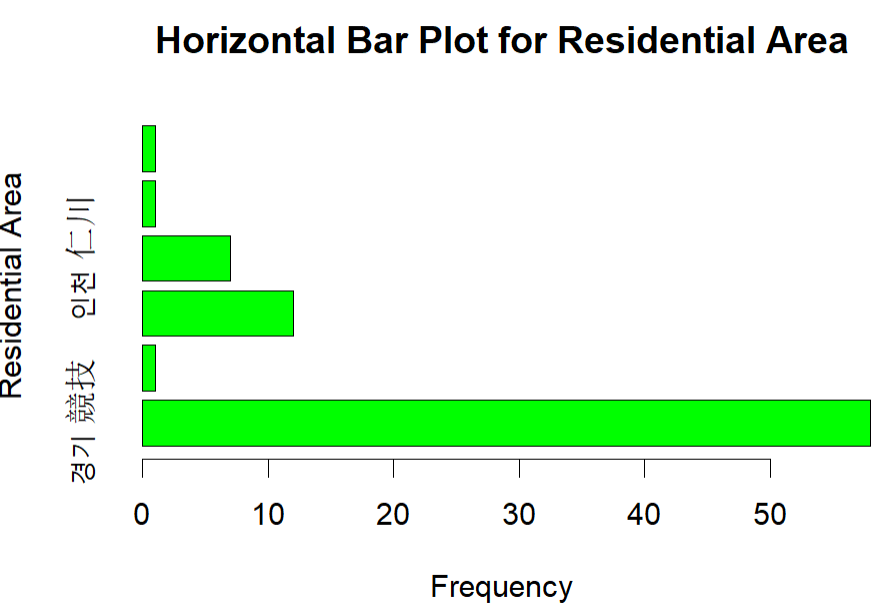
# step 5





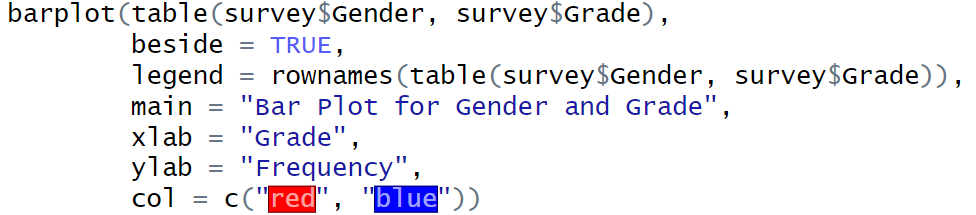
# step 6

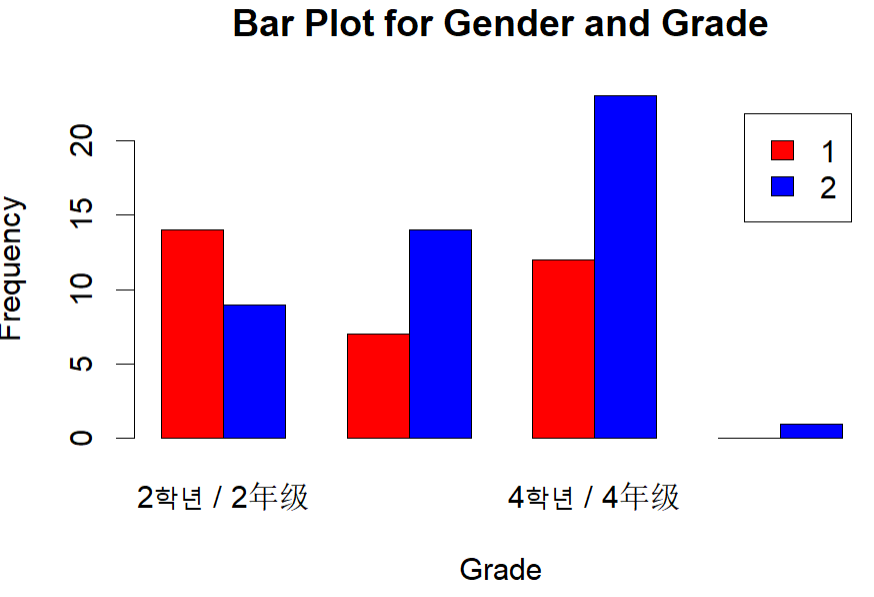




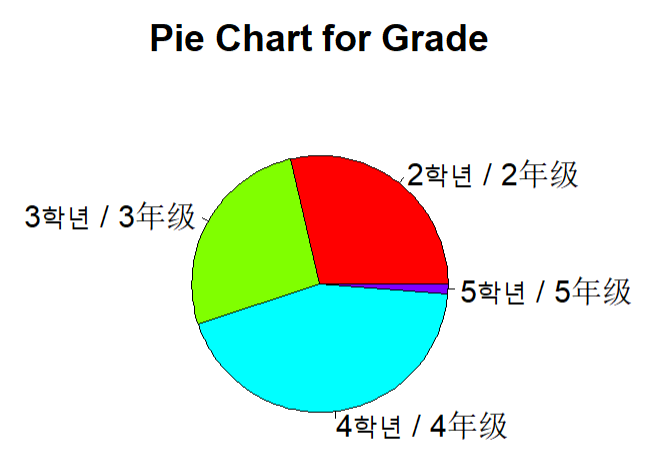
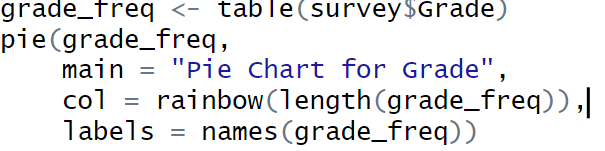
# step 7

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

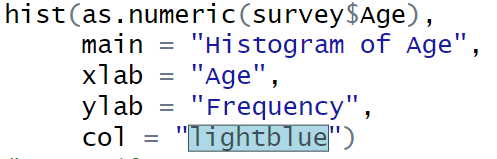


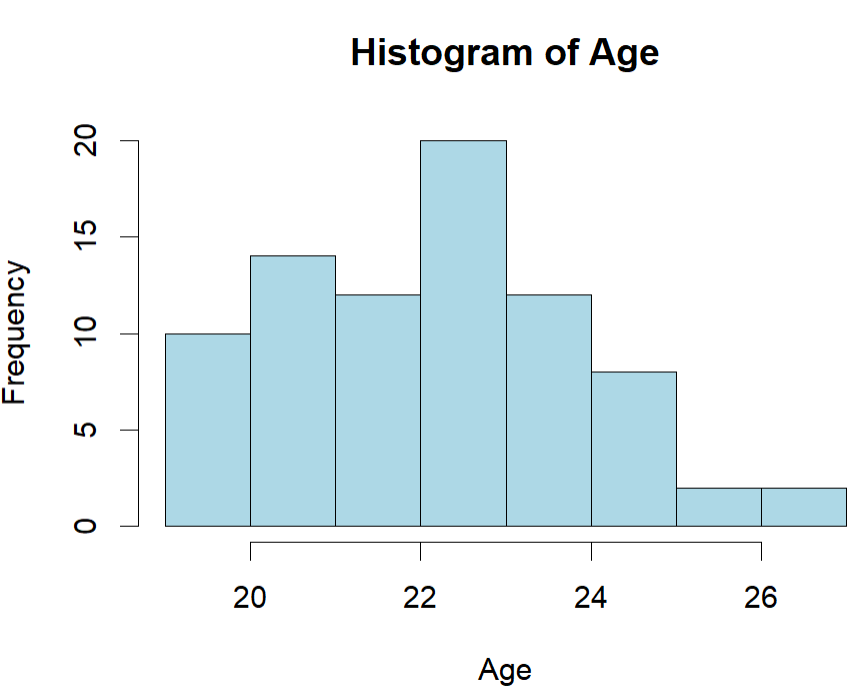


# step 8

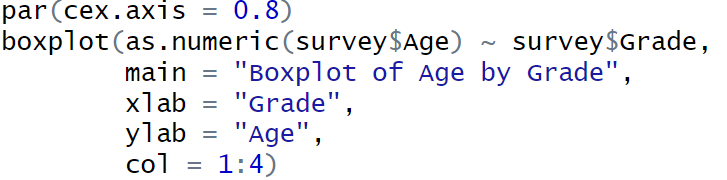


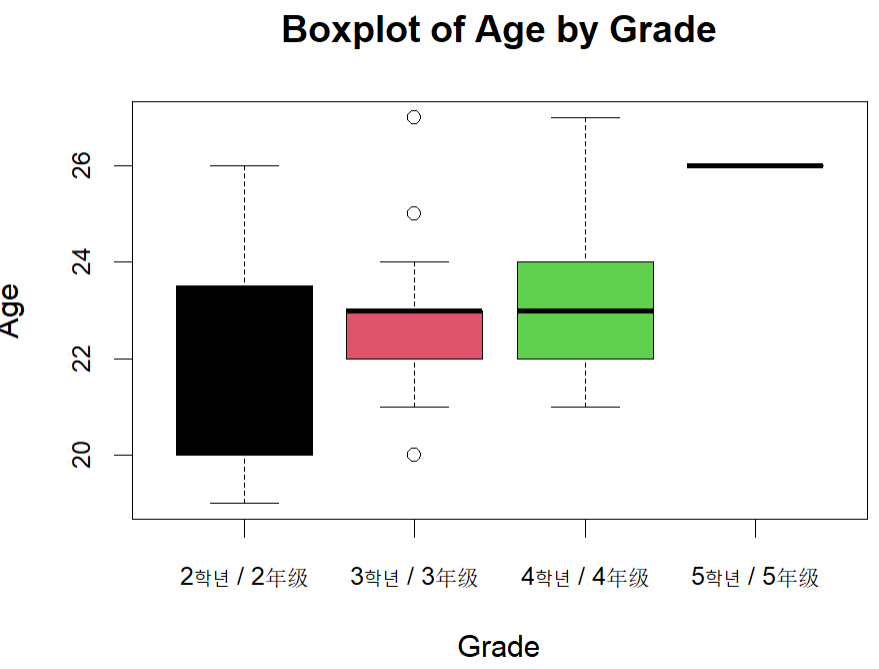
# step 9





# step 10





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

3학년 학생의 연령은 22~23세에 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 22세에서 24세 사이에 밀집되어 있으며 사분위수 범위가 약간 더 넓다는 것은 연령 분포가 3학년보다 약간 더 분산되어 있음을 나타냅니다.

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

# step 11

