통계패키지 활용 자료분석 (2020 년 2 학기)

담 당 교 수 : 김 태 수

강좌번호	100982-31001	본인의 과제 자체 평가	80 점

과제명 : 중간고사

이름	김태형

제 출 일	2020 년 10 월 22 일
학 과	기초교육학부(교류학생)
학 번	20180088

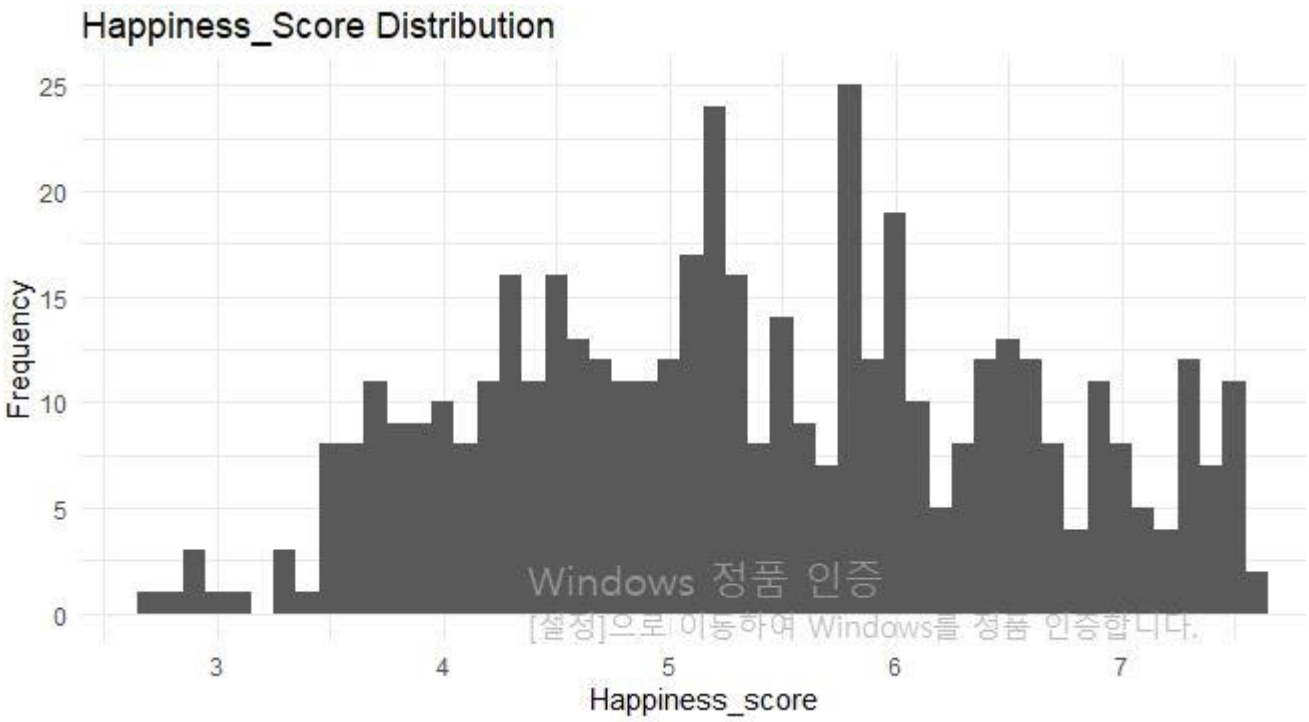
목차

1. 자료설명

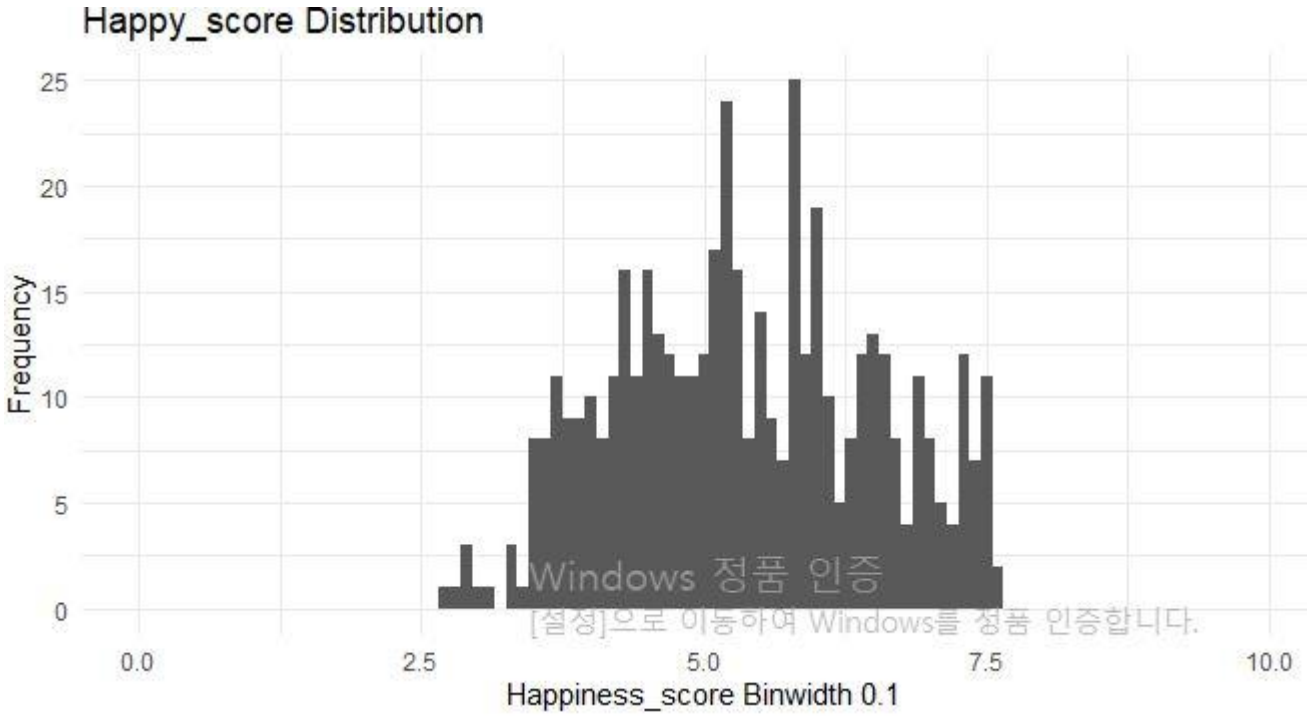
제시된 데이터는 2015 년에서 2017 년까지 세계의 행복과 관련된 수치를 보여주는 데이터이다. 총 10 개의 행과 243 개의 열로 구성되어 있다.

2. 시각화

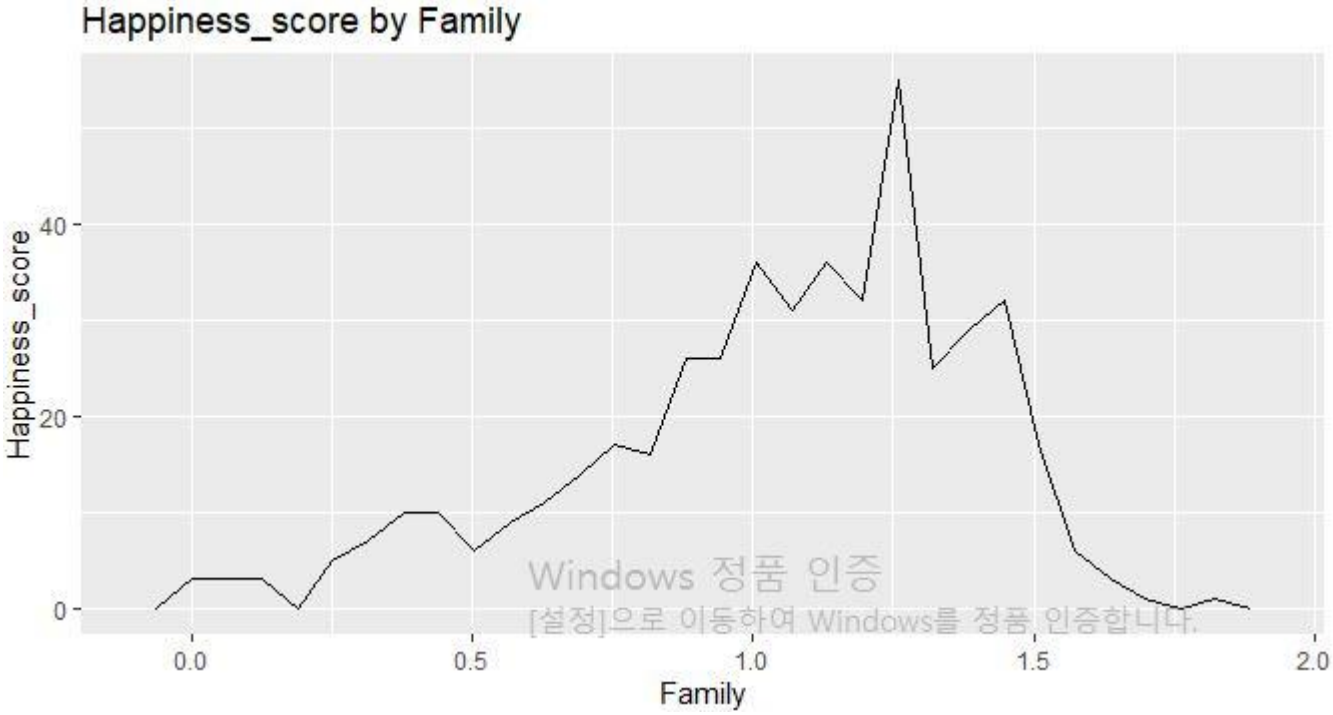
```
ggplot(data=base01) +
  geom_histogram(binwidth=0.1, aes(x=base01$Happiness.Score)) +
  ggtitle("Happiness_Score Distribution") +
  xlab("Happiness_score") + ylab("Frequency") + theme_minimal()
```



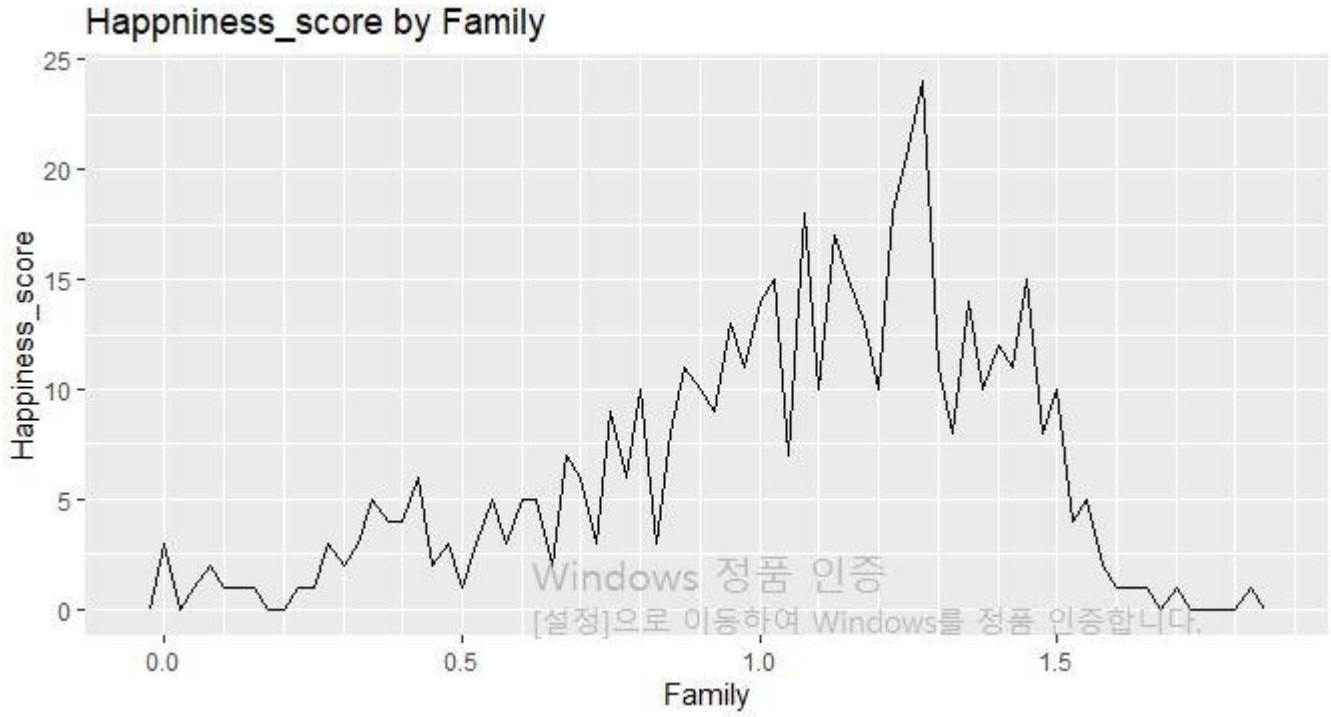
```
ggplot(data=base01) +
  geom_histogram(binwidth=0.1, aes(x=base01$Happiness.Score)) +
  ggtitle("Happy_score Distribution") +
  xlab("Happiness_score Binwidth 0.1") +
  ylab("Frequency") + theme_minimal() + xlim(0,10)
```



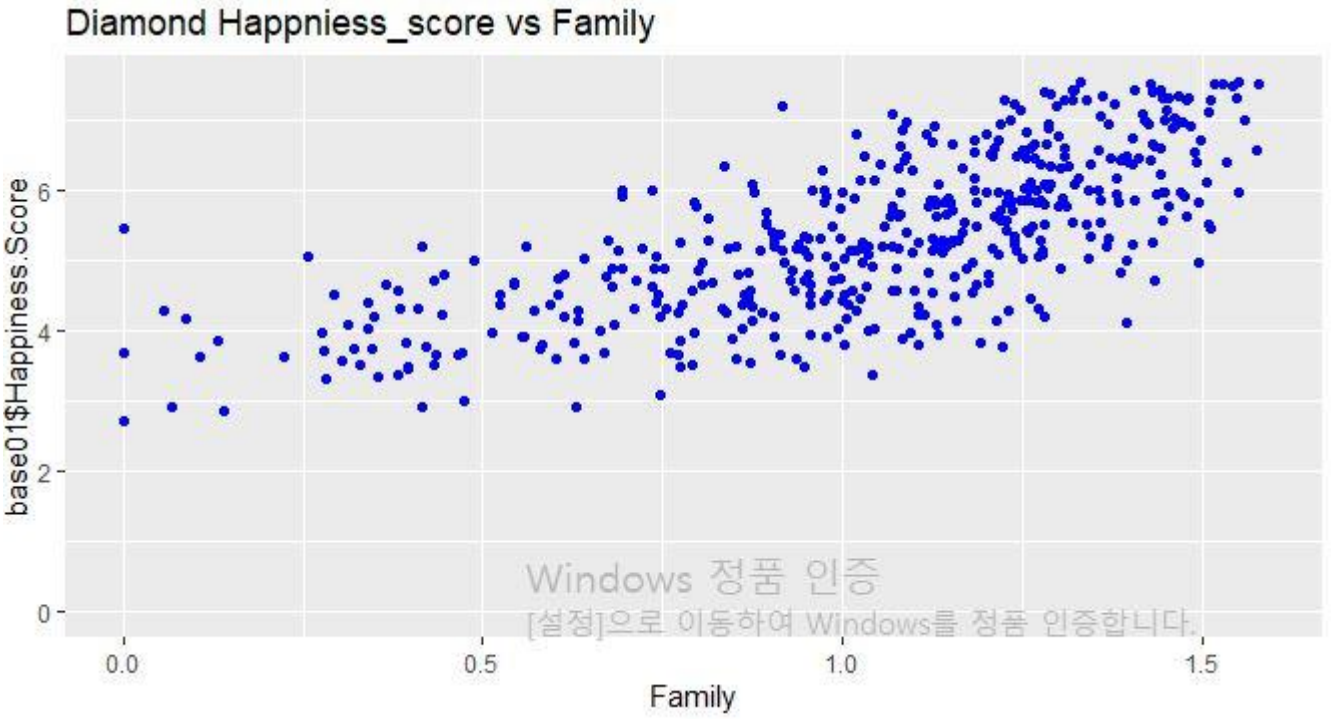

```
ggplot(data=base01, aes(x=Family)) + geom_freqpoly() +
  ggtitle("Happiness_score by Family") + xlab("Family") + ylab("Happiness_score")
```



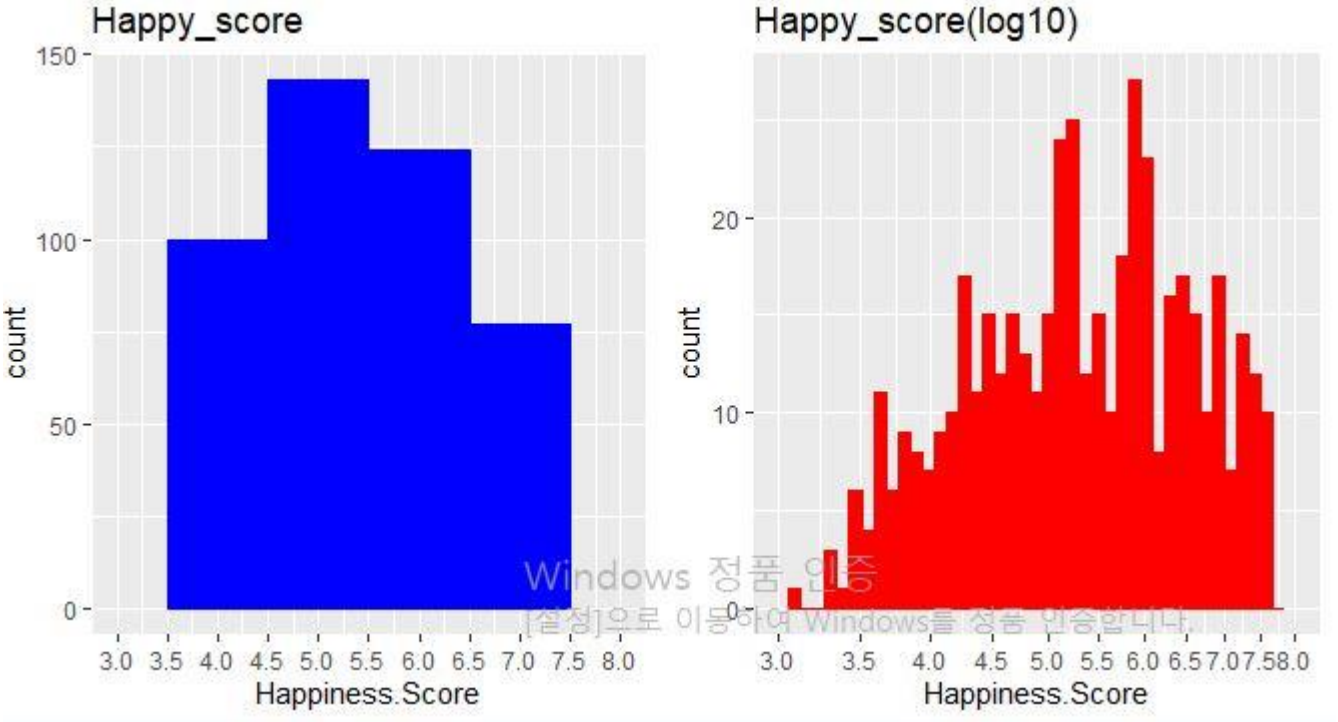
```
ggplot(data=base01, aes(x=Family)) + geom_freqpoly(binwidth = 0.025) +
  ggtitle("Happniness_score by Family") + xlab("Family") +
  ylab("Happiness_score") + scale_x_continuous(minor_breaks = seq(0, 5.5, 0.1))
```



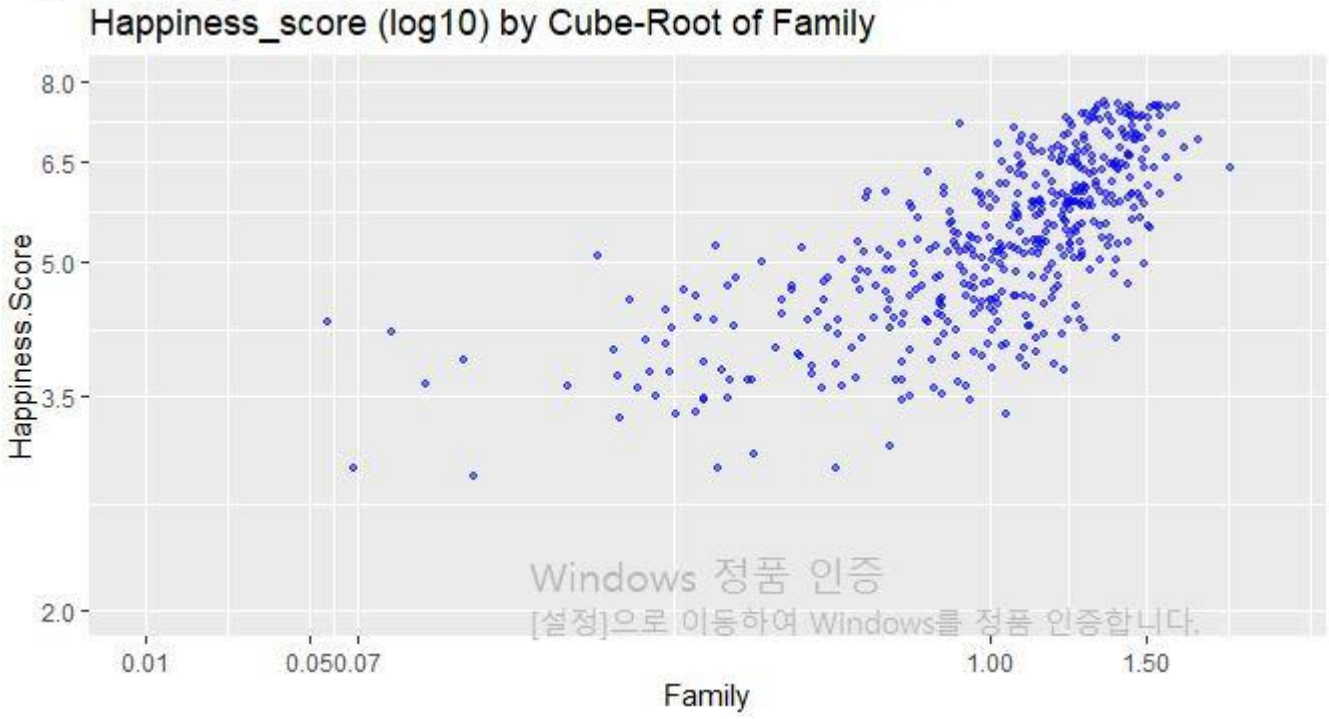

```
ggplot(base01,aes(x=Family,y=base01$Happiness.Score))+
  geom_point(color='blue',fill='blue')+
  xlim(0,quantile(base01$Family,0.99))+
  ylim(0,quantile(base01$Happiness.Score,0.99))+
  ggtitle('Diamond Happniess_score vs Family')
```



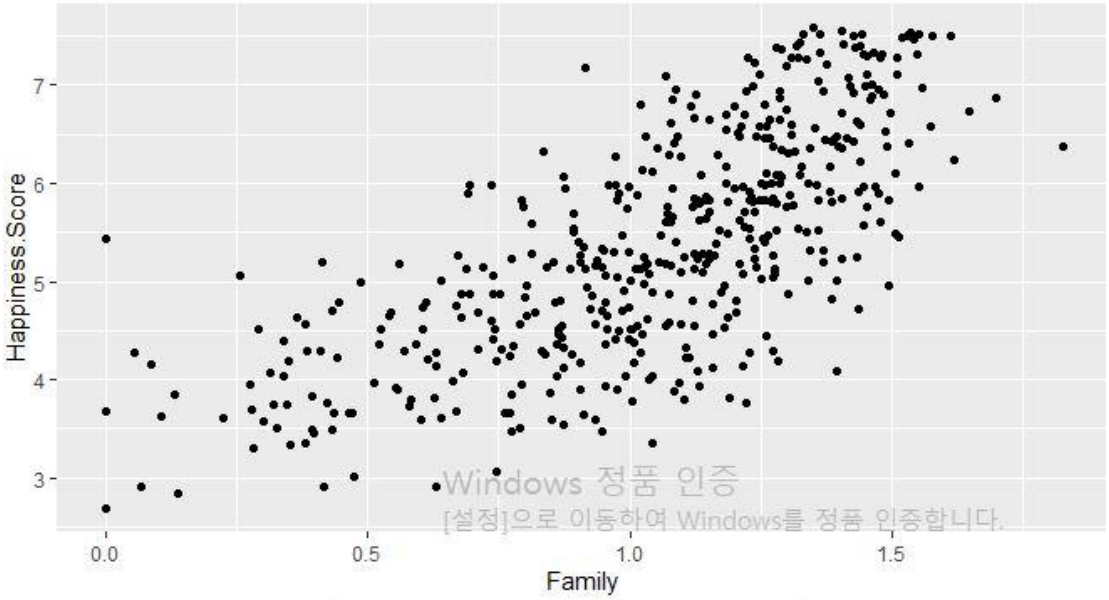
```
library(gridExtra)
plot1 <- ggplot(base01,aes(x=Happiness.Score))+
  geom_histogram(color='blue',fill = 'blue',binwidth=1)+
  scale_x_continuous(breaks=seq(3,8,0.5),limit=c(3,8))+
  ggtitle('Happy_score')
plot2 <- ggplot(base01,aes(x=Happiness.Score))+
  geom_histogram(color='red',fill='red',binwidth=0.01)+
  scale_x_log10(breaks=seq(3,8,0.5),limit=c(3,8))+
  ggtitle('Happy_score(log10)')
grid.arrange(plot1,plot2,ncol=2)
```

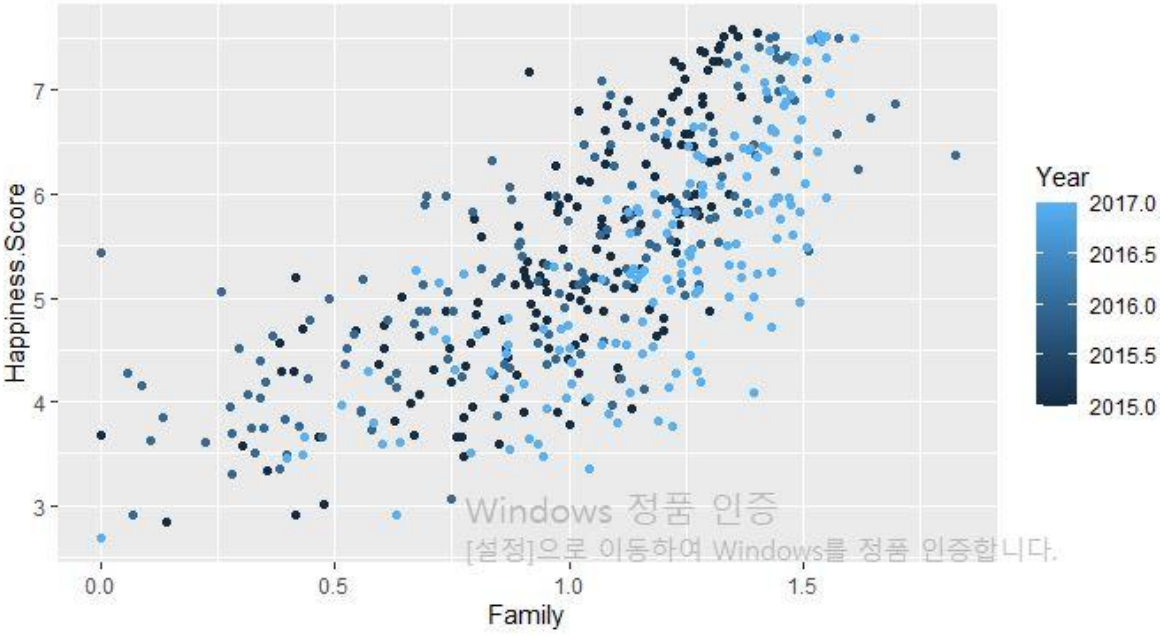


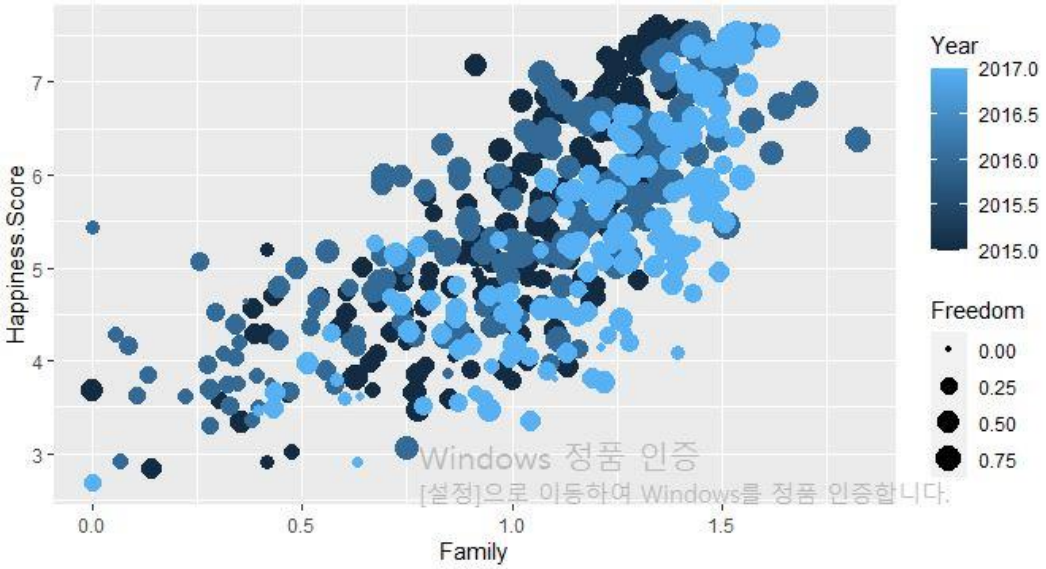
```r
library(scales)
cuberoot_trans = function() trans_new('cuberoot',
                                      transform = function(x) x^(1/3),
                                      inverse = function(x) x^3)

## Use the cuberoot_trans function
library(ggplot2)
ggplot(aes(Family, Happiness.Score), data = base01) +
  geom_point(color='blue',fill='blue',alpha=1/2,size=1,position = 'jitter') +
  scale_x_continuous(trans = cuberoot_trans(), limits = c(0.01, 2),
                     breaks = c(0.01, 0.05, 0.07, 1, 1.5)) +
  scale_y_continuous(trans = log10_trans(), limits = c(2, 8),
                     breaks = c(2, 3.5, 5, 6.5, 8)) +
  ggtitle('Happiness_score (log10) by Cube-Root of Family')
```

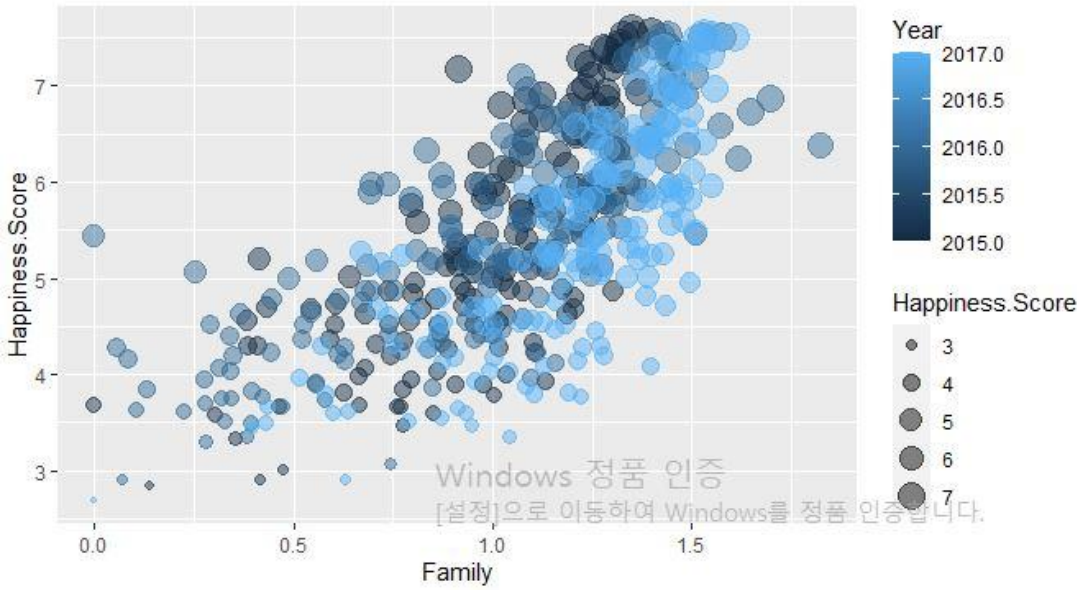


```r
##가로 축은 Family로, 세로축은 Happiness.Score로 설정, geom_point로 설정.
ggplot(base01, aes(Family, Happiness.Score))  + geom_point()

##색상추가
ggplot(base01, aes(Family, Happiness.Score))  + geom_point(aes(colour = Year))

##point의 크기는 꽃잎의  넓이(Happiness.Rank)에 따라 설정
ggplot(base01, aes(Family, Happiness.Score))  + geom_point(aes(colour = Year, size=Freedom))


##중복되어있는 점(겹쳐있는 점)을 표현
ggplot(base01, aes(Family, Happiness.Score))  + geom_point(aes(colour = Year, size=Happiness.Score), alpha=I(0.47))
```

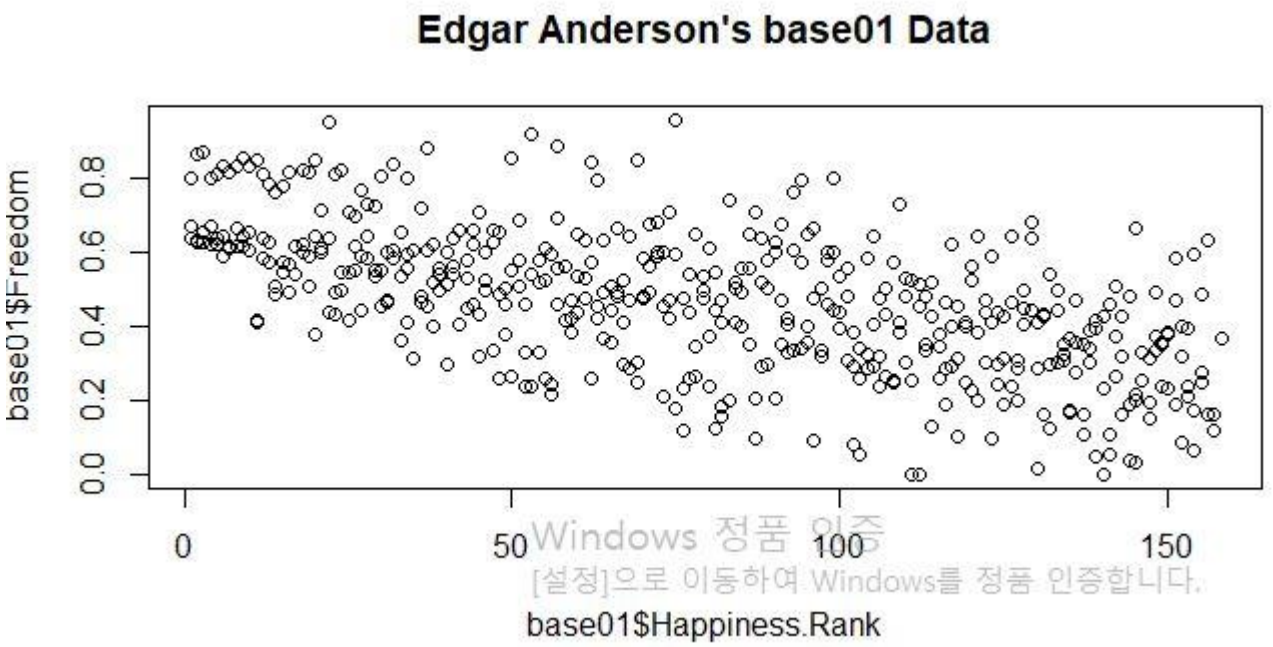


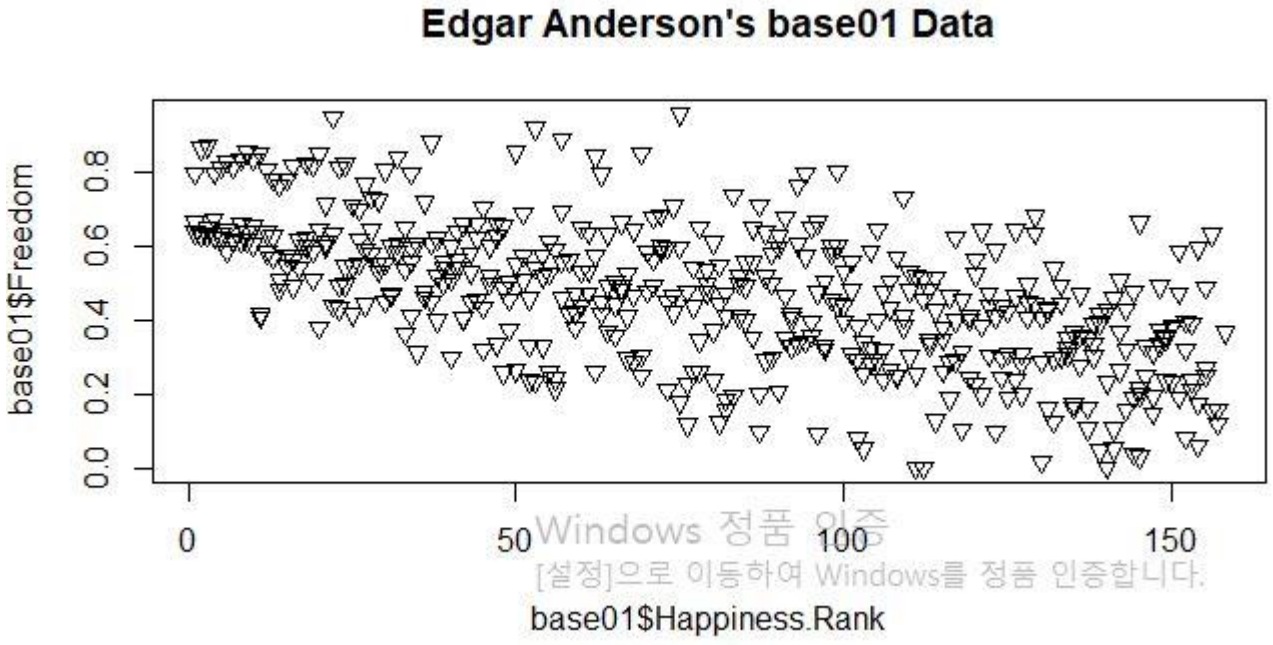



Happiness.Score
Family
Year
2017.0
2016.5
2016.0
2015.5
2015.0
Freedom
0.00
0.25
0.50
0.75
Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증합니다.


Happiness.Score
Family
Year
2017.0
2016.5
2016.0
2015.5
2015.0
Happiness.Score
3
4
5
6
7
Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증합니다.

```r
##Simple Scatter Plots
plot(base01$Happiness.Rank, base01$Freedom, main="Edgar Anderson's base01 Data")


plot(base01$Happiness.Rank, base01$Freedom, pch=c(21,22,23)[unclass(base01$Year)],
     main="Edgar Anderson's base01 Data")

plot(base01$Happiness.Rank, base01$Freedom, pch=25, bg=c("red","green3","blue")
     [unclass(base01$Year)], main="Edgar Anderson's base01 Data")
```





3. 결론

주어진 데이터를 시각화한 결과, 행복 점수(Happiness Score)는 가족(Family)와 양의 상관관계를 가지고 있음을 알 수 있다. 또한 행복 점수의 분포는 전반적으로 고르게 분포되어 있으나, 중위값에 대한 집중도가 좀 더 높다고 볼 수 있다.