보고서 - 데이터 탐색

학과:컴퓨터 메카트로닉스 공학부

학번:2017101281

이름:진윤화

##### 1.평균 연령과 2017년과 2016년의 평균 카드사용액과 평균 카드 사용건수를 구하십시오. (사용한 함수가 생성한 결과물을 화면 캡처해서 삽입하십시오)

readxl\_data <- read\_excel("sample1.xlsx")

head(readxl\_data)

tail(readxl\_data)

dim(readxl\_data)

str(readxl\_data)

readxl\_data$ID <- as.character(readxl\_data$ID)

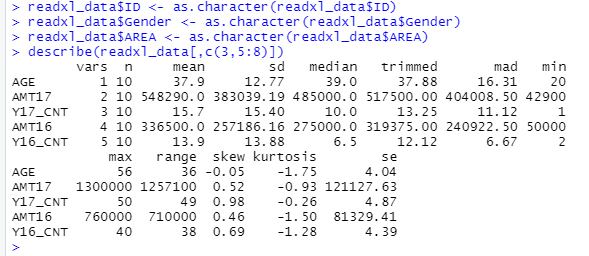
readxl\_data$Gender <- as.character(readxl\_data$Gender)

readxl\_data$AREA <- as.character(readxl\_data$AREA)

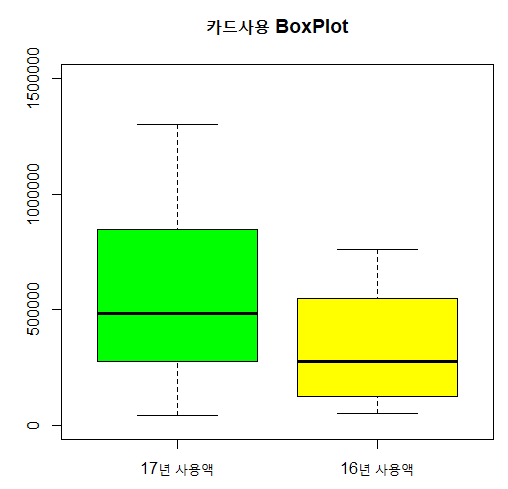
str(readxl\_data)

summary(readxl\_data)

describe(readxl\_data[,c(3,5:8)])

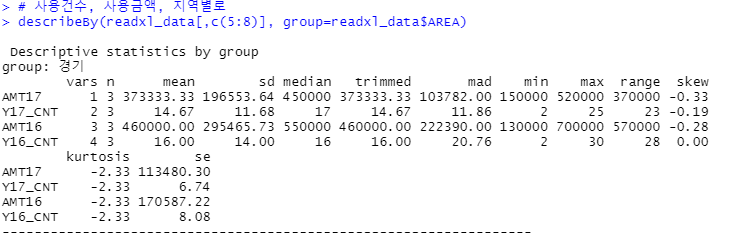


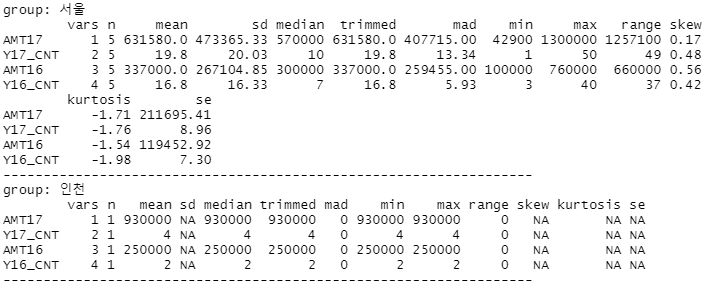
##### 2. 박스 플롯을 이용하여 2016년 카드 사용액과 2017년 카드 사용액을 도식화한 후, 차이를 간략히 설명하십시오 (Boxplot 결과물을 캡처하여 삽입하십시오)

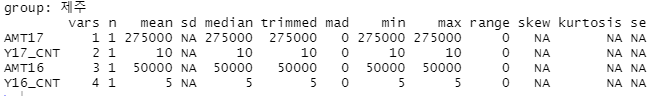


2017년 사용액이 2016년 사용액보다 높은 것을 알 수 있다.

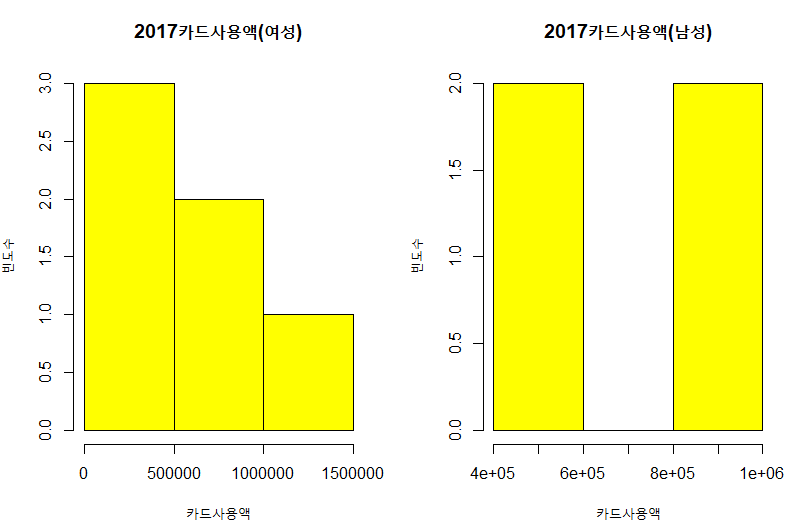
##### 3.2016년 카드 사용액과 사용횟수 2017년 카드 사용액과 사용횟수를 지역별로 기술 통계값을 산출하십시오





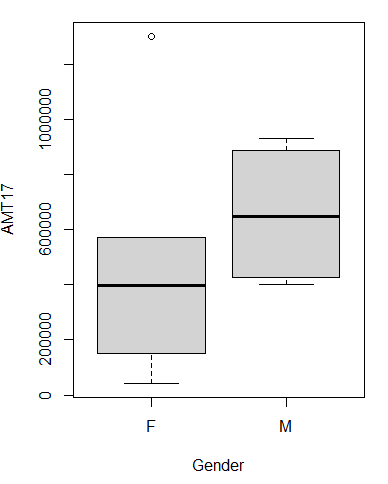


##### 4. 2017년 카드 사용액을 성별별로 히스트그램을 그리시오. 계급의 수나, 히스토그램의 색상, 재목 등은 가독성이나 이해력을 고민하여 정의하십시오. 그리고 간단히 차이를 기술하십시오.



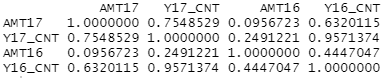
##### . 4.1. 가능한 경우, 다른 변수에 대하여 히스토그램을 그려보십시오

##### 5. 2017년 카드 사용액을 성별별로 박스플롯을 그리시오. 가독성이나 이해력을 높이기 위해 필요한 박스플랏 옵션을 이용하십시오



##### 5.1. 가능한 경우, 다른 변수에 대하여 박스플롯을 그려보십시오.

##### 6. 2016년, 2017년 카드 사용건수 사용금액의 상관관계를 구하시오



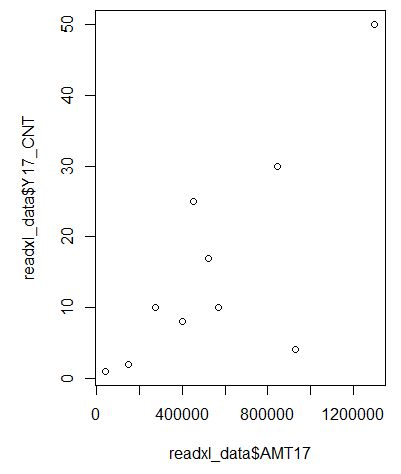
16, 17년 카드 사용금액과는 상관도가 없음

그러나 16, 17년 사용 횟수와는 상관도가 높음

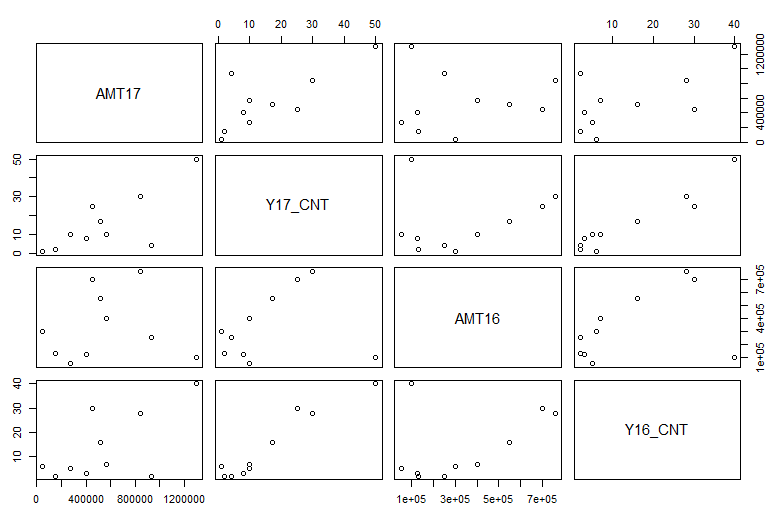
따라서 16년에 카드 사용건수가 높을수록 17년 사용건수는 높다.

사용횟수와 사용금액은 상관도가 높음.

##### 7. 2017년 카드사용액과 카드사용건수의 산점도를 그리시고, 플롯 옵션을 이용하여 가독성 있게 그려 보시오.



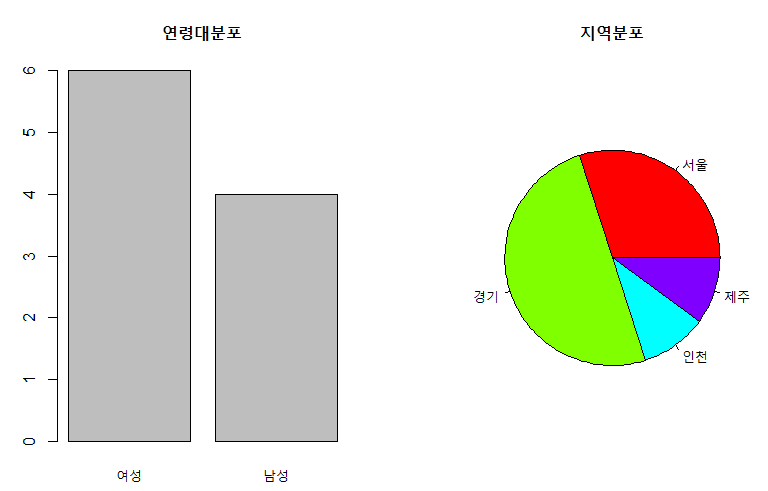
##### 8. 2016년도, 2017년도 카드 사용건수와 사용금액 간의 산점도를 그리시오



##### 9. 6번과 8번을 토대로, 카드사용건과 사용금액 간의 관계를 간략히 기술하십시오

카드 사용 건이 높아질수록, 사용금액이 높아지는 관계가 있다.

##### 10. 성별과 지역의 도수(빈도)와 상대도수를 구하시고, bar chart와 pie chart를 그리시오



##### 11. 성별별로, 2년간의 카드사용액 총합 (AMT)의 평균값, 합계, 최소값, 최대값, 카드 사용자수(빈도)를 알아보자

