

5강. 이변량 자료의 시각화 2

◆ 담당교수 : 허명희

들어가기

■ 주요용어

용어	해설
모자이크 플롯	교차표의 시각화 기법으로 다차원 칸을 도수(count)에 비례하는 면적의 사각형으로 표현한 큰 사각 도형이다.
나무 지도	범주의 크기를 사각형의 너비로, 범주의 밀도 특성을 사각형의 컬러로 표현한 사각 도형으로 포트폴리오 자료의 시각화에 유용하다.

연습문제

1. 모자이크 플롯을 위한 R의 함수는 무엇인가?

정답 : mosaicplot()

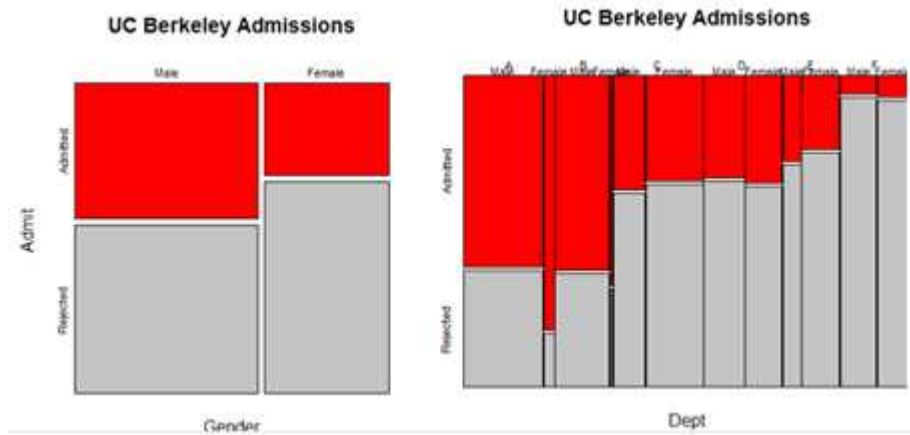
2. 나무 지도를 그리기 위한 R의 함수는 무엇인가?

정답 : treemap 패키지의 treemap()

3. 버클리 입학자료(UCBAdmissions)에서 Berkeley 대학원 입학에서의 성차별 문제에 대한 자료 분석적 답을 제시하라.

정답 : data(UCBAdmissions)
str(UCBAdmissions)
windows(height=5, width=4.6)
mosaicplot(~Gender+Admit, data=UCBAdmissions,
color=c("red","grey"), main = "UC Berkeley Admissions")
windows(height=5, width=6)

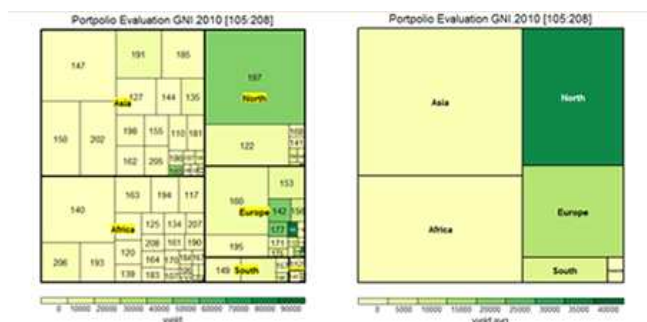
```
mosaicplot(~Dept+Gender+Admit, data=UCBAdmissions,
           color=c("red","grey"), dir=c("v","v","h"), off=1,
           main = "UC Berkeley Admissions")
```



4. GNI.2010 자료에서 뒷부분 104개 개체에 대한 나무 그림을 그려라. 첫부분 104개 개체에 대한 나무 그림인 그림 15그림 16과는 어떻게 다른가?

정답 : library(treemap)

```
GNI.2010 <- read.table("GNI-2010.txt", header=T)[105:208,]
str(GNI.2010)
windows(height=8, width=7)
treemap(GNI.2010,index=c("sector","item"),vSize="principal",
        vColor="yield",type="value",bg.labels="yellow",
        title="Portpolio Evaluation GNI.2010 [105:208]")
GNI.2010$yield.total <- GNI.2010$principal*as.numeric(GNI.2010$yield)
GNI.2010.a <- aggregate(GNI.2010[,3:5],by=list(GNI.2010$sector),sum)
GNI.2010.a$yield.avg <- GNI.2010.a$yield.total/GNI.2010.a$principal
windows(height=8, width=7)
treemap(GNI.2010.a,index=c("Group.1"),vSize="principal",
        vColor="yield.avg",type="value",bg.labels="yellow",
        title="Portpolio Evaluation GNI.2010 [105:208]")
```



정리하기

1. 모자이크 플롯은 교차표를 시각화한다. 범주형 변수들의 결합 칸을 빈도에 비례하는 크기의 모자이크 타일로 표현한다.
2. 나무 지도(tree map)는 타일을 계층적으로 배열하여 붙인 그림이다. 타일의 크기는 개체의 규모를, 컬러는 밀도적 속성을 나타낸다. 대표적인 적용 예는 금융 포트폴리오 자료인데 나무 그림에서 영역별·항목별 투자 규모와 투자 수익률을 한 눈에 파악할 수 있다.

참고자료

1. 허명희 (2010). R을 활용한 탐색적 자료분석 (1판 3쇄), 자유아카데미.
2. 허명희 (2014). 데이터 시각화, 자유아카데미.