**Practice 1 ( until 2016-10-01 09:00 )**

**1. Crime.csv는 2005년 미국의 범죄율 데이터로 범죄 유형별 발생건을 인구 100,000명 중의 발생 비율로 표시하였다. 살인, 강도, 폭행, 절도 등 총 7가지 범죄를 포함하는데 이 중 살인(murder)와 절도(burglary) 사이의 관계를 살피려고 한다.**

1. 두 변수 사이의 산점도를 단변량 분포와 함께 그리시오. 상관계수도 함께 살피시오.

|  |
| --- |
| Crime <- read.csv("../0910/crime.csv",stringsAsFactors=F)  pairs.panels(Crime[,-1]) # Cor : 0.28 |

1. 위를 통해 이상점 존재여부를 판단하고 존재한다면 해당 주를 확인하고 제거하시오. 제거 후 변수들 사이의 관계가 어떻게 변화하는지 살피시오.

|  |
| --- |
| plot(Crime$burglary, Crime$murder,xlim=c(0,800))  identify(Crime$burglary,Crime$murder,labels=Crime$state)   * # District of Columbia 가 이상치. |

|  |
| --- |
| # United States 자체가 전체 통합 데이터 이기 때문에 같이 제거  out=match(c("United States","District of Columbia"),Crime$state)  clr=rep(1,dim(Crime)[1])  clr[out]=2  pairs(Crime[,-1],col=clr,pch=clr)    # 이상치 제거 이후 Cor 확인  # Murder Burglary : 0.62  Crime2<-Crime[-out,] *# 이상치 제거*  pairs.panels(Crime2[,-1]) *# cor: 0.62* |

1. 살인, 절도와 인구(population)의 관계를 함께 관찰하기 위해 bubble plot을 그리고 관찰한 사실을 기술하시오.

|  |
| --- |
| with(Crime2, symbols(murder,burglary,circles = population))   * 인구가 많으면 살인과 강도가 많이 발생하나 무조건 적이지는 않다. * 인구가 가장 많다고 해서 살인과 강도가 가장 많이 일어나는 것은 아니다. |

1. 7가지 범죄의 발생 건수를 heatmap, 별그림, 나이팅게일 차트로로 표현하고 범죄 발생 특징 간의 패턴이 비슷한 주들이 있는지 살피시오.

|  |
| --- |
| plot(Crime2$forcible\_rape, Crime2$murder)  identify(Crime2$forcible\_rape,Crime2$murder,labels=Crime2$state   * Burglary와 Murder 관계 외 forcible\_rape와 murder 관계에서의 이상점인 Alasaka 삭제     out=match(c("United States","Alaska","District of Columbia"),Crime$state)  Crime3 <- Crime[-out,]  pairs.panels(Crime3[,-1])    **library**(RColorBrewer)  TmpCrime <- Crime3  rownames(TmpCrime) <- Crime3[,1]  TmpCrime <- TmpCrime[,-1]  TmpCrime <- as.matrix(TmpCrime)  heatmap(TmpCrime[,-8], scale="column", col=brewer.pal(9,"Blues"))  # 나누는 기준은 개인마다 차이가 있으므로 개인의 판단에 따라 나누면 됨.    Star Plot  stars(TmpCrime[,-8],cex=0.7, key.loc=c(21,5))    Nightingale’s Chart  stars(TmpCrime[,-8],cex=0.7, key.loc=c(21,5),draw.segments = TRUE) |