

08. 데이터 시각화 # src: 7_R/08_데이터시각화 참고

1) 시각화: 데이터 ⇒ 데이터 가공 ⇒ 시각화 ⇒ 꾸미기 #꾸미기: R언어X, 이미지 편집도구 사용

(1) 데이터 가공: 사용하기 쉬운 형태로 데이터 변환(base, stats, plyr, reshape, reshape2 등 이용)

(2) 시각화: graphics, ggplot, ggplot2 등의 패키지 사용하여 원 데이터나 분석된 결과 시각화

시각화 종류	상세
산점도(Scatter plot)	색, 모양, 크기 추가, 산점도 행렬
선 그래프(Line graph)	line: 방향, path: 무방향, Time series plot(시계열 그래프)
히스토그램(Histogram)	연속형 변수의 빈도수 분포, 단일 변수 차트
밀도 그래프(Density)	밀도 값으로 그린 선 그래프
막대 그래프(Bar chart)	이산형 변수 빈도수 분포표, 단일 변수 차트, 도수 분포표, 누적 막대그래프
박스 그래프(Box plot)	사분위수와 이상값 표시
모자이크 플롯(Mosaic plot)	이산형 데이터의 다차원 도수 분포표
파이 차트(Pie chart)	이산형 변수의 빈도수 분포, 단일 변수 차트

2) **고수준 그래프 함수**: 그래프 함수 호출시마다 초기화 후 새로운 그래프를 다시 그림

plot(), barplot(), boxplot(), hist(), curve(); 수식을 그래프로, qqnorm(); 분위수-분위수 그래프

- par(); 그래프를 조정하거나 그래프 창의 특성 지정(선 굵기, 종류, 문자 크기, 글꼴, 색상 등)
- plot(x, y, ...); ...에는 type; 그래프 유형 main; 그래프 제목, xlab; x축 제목, ylab, asp; y/x 종횡비

3) **저수준 그래프 함수**: 새 그래프 생성X, 이미 그려진 그래프에 점, 선, 텍스트, 장식 등 추가

points(), lines(), abline(); 직선, polygon(); 닫힌 다각형, text(), segments(); 선분, 두 점을 잇는 직선

4) 그래프 파라미터

(1) byt 그래프의 박스타입 설정 `o l t c u` 문자 모양대로 테두리 만들어짐 n: 박스 테두리無

(2) cex 텍스트 또는 기호 크기 설정 font or ps 글꼴 설정 (1:normal, 2: italics, 3:bold, ...)

(3) col 색 설정(col=숫자 or col="색") las 눈금 라벨 (0:축과 평행, 1: 수평, 2: 축에 수직, 3: 수직)

(4) mar 그래프 여백(하-좌-상-우 순) pch 점의 표시 유형 설정(하기 번호 및 모양 참조)



(5) new 기본값(F), T: 고수준 그래프 함수 이용하더라도 이전 그래프 삭제하지 않고 추가

5) **ggplot2**

(1) ggplot2 시각화 레이어: ①배경 설정(축) ⇒ ②그래프 추가(점, 막대, 선 등) ⇒ ③설정 추가

(2) 기하객체(geom); geom_point(); 산점도, geom_smooth(); 회귀선, geom_line(), geom_density(), geom_histogram(), geom_bar(); 막대 그래프, 도수 분포표, geom_pointrange(): 값과 범위 표시

(3) 미적 속성(aes); x, y:좌표 값, color, shape, size, alpha: 투명도, fill: 채워 넣는 색