

6. R

6.1 데이터 시각화 단계

1) 데이터 가공

- 데이터를 R Language에서 사용하기 쉬운 형태로 변환
- 주로 base, stats, plyr, reshape, reshape2 등의 패키지를 활용

2) 시각화

- R Language를 사용하여 원 데이터를 시각화 하거나 분석된 결과를 시각화 한다.
- graphics, ggplot, ggplot2 등의 패키지를 기본적으로 사용한다.
- 생성된 시각화 이미지는 jpg, png, pdf 형태로 저장할 수 있다.

3) 꾸미기

- 시각화 결과를 보고 하거나 외부로 배포할 경우, 의미를 명확하게 전달하기 위해서 인포그래픽스 작업을 진행한다.
- 꾸미기 작업은 R Language가 아닌 외부 전문 이미지 편집 도구를 이용한다.

6.2 시각화의 종류

종류	유사어	상세
산점도	점 그래프 Scatter plot	- 색, 모양, 크기 추가 - 산점도 행렬(Scatter plot matrix)
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	- 값이 아니라 밀도 값으로 그린 선 그래프
막대 그래프	Bar chart	- 이산형 변수의 빈도수 분포표, 단일 변수 차트 - 도수분포표(Frequency table) 또는 누적 막대 그래프
박스 그래프	Boxplot	- 사분위수와 이상값 표시
모자이크 플롯	Mosaic plot	- 이산형 데이터의 다차원 도수 분포표
파이 차트	Pie chart	- 이산형 변수의 빈도수 분포, 단일 변수 차트