

1강. 데이터 시각화란 무엇인가?

◆ 담당교수 : 이태림

들어가기

■ 학습개요

데이터 시각화는 데이터에 내재된 통계적 정보를 그림으로 드러내는 것이다. 효과적인 시각화를 위해서는 데이터 성격에 맞는 최적의 그래프를 선택하고 이를 뒷받침하는 소프트웨어를 활용할 필요가 있다. 이 강의에서는 데이터 시각화의 기본을 다룬다. 오픈소스의 통계계산 소프트웨어인 R을 활용한다. R의 컬러 체계와 그래프 기본 요소를 설명하고 시각화 과정의 보기를 제시한다.

■ 학습목표

1	데이터 시각화의 개념과 배경을 이해할 수 있다.
2	R 그래프의 컬러 체계와 여러 기본 요소를 학습할 수 있다.
3	시각화 과정에 익숙해질 수 있다.

■ 주요용어

용어	해설
데이터 시각화	데이터에 내재된 통계적 정보를 그림의 형태로 드러내는 것이다.
소프트웨어 R	오픈소스의 통계계산 소프트웨어로서 시각화에 강점이 있다.
(R, G, B) 컬러	“red”, “green”, “blue”의 조합으로 모든 컬러를 표현한다.
점, 선, 다각형	그래프의 기본 요소이다.

연습문제

1. 데이터 시각화란 무엇인가?

정답 : 데이터 시각화는 데이터에 내재된 통계적 정보를 그림의 형태로 드러내는 것이다.

2. 데이터 시각화 과정에서 스토리텔링의 역할은 무엇인가?

정답 : 스토리텔링은 그림의 핵심적 의미를 일반인에게 전달하는 커뮤니케이션이다.

3. R의 rainbow 함수를 써서 아래와 같이 무지개의 12개 색으로 구분되는 파이 차트(도넛 차트)를 만들어 제시하라.



```
정답 : pie(rep(1,12),col=rainbow(12),border="white",clockwise=TRUE,
          labels="")
      par(new=T)
      pie(rep(1,1),col="white",radius=0.5,border="white",labels="")
```

정리하기

1. 데이터는 다수 개체들에 대해 한 묶음의 특성이 담긴 원재료이다. 통계적 정보는 데이터의 전체 모습이라 하겠는데 특히 큰 규모의 데이터는 그것을 좀처럼 드러내지 않는다. 데이터 시각화의 목표는 데이터의 통계적 정보를 그림의 형태로 드러내는 데 있다.

2. 데이터 시각화의 과정

- ① 데이터의 이해
- ② 목표 설정
- ③ 그래프 선정
- ④ 그래프 구현

⑤ 스토리텔링

3. R은 오픈소스의 통계계산 언어이자 소프트웨어이다. 시각화에 강점이 있다.
4. R의 컬러는 (R,G,B) 코드 조합으로 구성되는데 R,G,B (red, green, blue)는 각각 256개 수준을 갖는다. 따라서 표현 가능한 컬러의 총 수는 1,600만개이다. 컬러 함수에는 rainbow, heat.colors, terrain.colors, topo.colors, cm.colors 등이 있다. R의 컬러 패키지로는 RColorBrewer 등이 있다.
5. R 그래프의 기본 요소는 점, 선, 사각형, 다각형 등으로 points, segments, rect, polygon 등의 함수로 표현된다.

	참고자료
--	-------------

1. 허명희 (2014). 데이터 시각화, 자유아카데미. (R 그래프의 방법들이 다루어졌다)