

샘플데이터 확인하기

diamonds 데이터 확인하기

- 데이터 핸들링에서 사용할 데이터는 ggplot2 패키지에서 제공하는 diamonds를 사용
- 53,940개의 행과 10개의 열(또는 변수)을 가지는 데이터 프레임을 가짐

변수명	변수설명
price	price in US dollars (\\$326–\\$18,823)
carat	weight of the diamond (0.2–5.01)
cut	quality of the cut (Fair, Good, Very Good, Premium, Ideal)
color	diamond colour, from J (worst) to D (best)
clarity	a measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best))
x	length in mm (0–10.74)
y	width in mm (0–58.9)
z	depth in mm (0–31.8)
depth	total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43–79)
table	width of top of diamond relative to widest point (43–95)

diamonds 데이터 확인하기

- 데이터 보기는 데이터의 값들이 어떻게 생겼는지를 확인하고 싶을 때에 사용함
- 데이터보기의 목적
- 데이터 전체를 보거나 일부를 볼 수 있음
- 외부 데이터를 읽어왔을 때에 잘 읽어왔는지를 확인하기 위함 데이터의 생성, 수정, 삭제 등이 제대로 되었는지를 확인하기 위함

diamonds 데이터 확인하기

- 데이터를 전체 는 두 가지 방법
 - 콘솔(console)에 보이도록 하는 방법
 - 데이터명을 실행
 - 최대 10,000개까지만 보이고 그 이상은 생략됨
 - 데이터 명 : diamonds
 - 에디터 화면에 보이도록 하는 것
 - View() 함수를 이용
 - 데이터를 모두 볼 수 있음
 - 각 변수를 기준으로 오름차순이나 내림차순으로 정렬하는 기능도 제공
 - 데이터 명 : View(diamonds)

diamonds 데이터 확인하기

- 데이터의 일부를 보는 두 가지 방법
 - head() 함수의 사용
 - 데이터의 끝 부분 중에서 일부를 보여줌
 - 데이터 명 : tail(diamonds)
 - tail() 함수의 사용
 - 데이터의 시작 부분 중에서 일부를 보여줌
 - 데이터 명 : head(diamonds)
- 두 함수 모두 해당 데이터에서 6개만 보여줌
- 보고 싶은 데이터의 개수를 조정하고 싶을 때 n이라는 argument 사용
 - head(diamonds, n=10)
 - tail(diamonds, n=10)