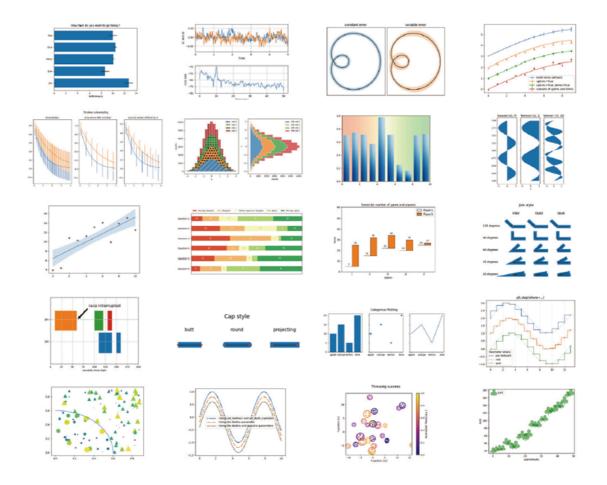
# 데이터 시각화(visualization)

## 1. 데이터 시각화란?

- 수많은 데이터를 눈으로 보고 의미를 파악하는 것은 어려움이 있음
- 데이터의 의미를 효과적으로 전달하기 위해 조형 요소를 이용하여 시각적으로 표현한 것
- graph, chart, plot, diagram

## 2. 데이터 시각화 장점

- 많은 양의 데이터를 시각적으로 요약
  - 시각화 요소인 도형의 형태, 크기, 위치, 색, 정보 등으로 시각적 패턴을 찾음
  - 정확한 전달을 위해 시각화 형태를 잘 선택해야 함
  - 데이터 시각화의 목적을 명확히 해야 함
- 더 정확한 데이터 분석 결과 도출



# 3. 시각화 기능

- 설명 기능
  - 데이터 시각화를 통해 전달하려는 메시지와 주요한 분석 결과를 설명하는 기능
  - 데이터로부터 유의미하거나 흥미로운 이야기와 분석을 명확하게 보여주어야 함

- 탐색 기능
  - 데이터에 숨겨져 있는 관계와 패턴을 찾기 위한 시각적 분석 기능
  - 데이터의 유의미하거나 흥미로운 요소를 이용자가 직접 탐색하는 것
- 표현 기능
  - 데이터에 대한 분석적인 통찰보다는 데이터를 활용한 개인 작품이나 예술적인 표현을 통해 감정적인 시선이나 이야기 전달, 공감을 불러일으키기 위한 기능

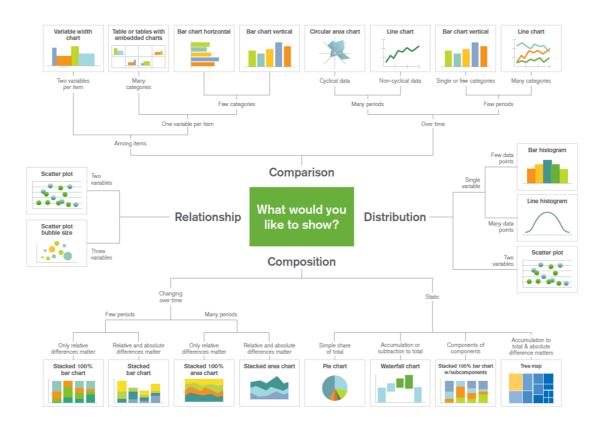
#### 4. 시각화 목적

- 정보 전달: 데이터의 진실을 간단하고 정확하게 전달
- 설득 : 데이터의 창의적이고 심미적 표현을 통해 데이터를 통해 전달하고자 하는 메시지에 대한 공감, 설득 등의 감정적 반응을 유도하는 추상적이고 예술적 측면의 목적

#### 5. 시각화 방법

• 데이터의 변수의 수, 종류, 분석 목적에 따라 시각화 도구가 달라짐

시간 시각화	분포 시각화	관계 시각화	비교 시각화	공간 시각화
<u>막대 그래프</u> 누적 <u>막대 그래프</u> 산점도 선 그래프 계단식 그래프 영역 차트	파이 차트 도우넛 차트 누적 <u>막대 그래프</u> <u>트리맵</u> 누적 영역 차트	산점도 산점도 행렬 버블 차트 히스토그램	<u>막대 그래프</u> <u>히트맵</u> 평행 좌표계 스타 차트 체르노프 페이스	지도 매핑



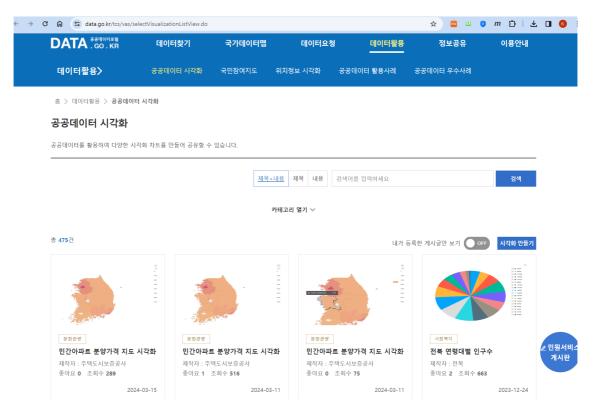
출처: https://blog.qlik.com/third-pillar-of-mapping-data-to-visualizations-usage

#### 6. 데이터 시각화 활용 사례

- 시각화 대시보드를 통한 데이터 시각화 공유
  - 주요한 데이터 지표를 모니터링하고 탐색시 유용
  - 대시보드(Dashboard) : 한 화면에서 다양한 정보를 중앙 집중적으로 관리 하고 찾을 수 있도록 하는 사용자 인터페이스 기능



• 공공부문, 민간 부문에서 보유하고 있는 데이터를 활용 및 시각화



https://www.data.go.kr/tcs/vas/selectVisualizationListView.do

#### 7. 파이썬 기반 시각화 라이브러리

- Matplotlib: 파이썬 표준 시각화 라이브러리. 가장 많이 사용. Pandas 시각화기반
- Seaborn : Matplotlib 기반에 스타일을 추가. 다양한 그래프 내장
- Folium : 지도 데이터를 이용하여 위치정보를 시각화. 자바스크립트 기반의 상호작용그래프 작성
- PyCharts : Baidu에서 시각화를 위해 만든 파이썬 버전의 라이브러리. 자바스트립트 기반
- Plotnine : R의 ggplot2에 기반하여 그래프 그려줌
- Plotly : 상호작용 그래프를 그려주는 라이브러리. R, Scalar, Python, Javascript 등에서 사용. 사용이 쉽고 세련된 도구
- Bokef : Matplotlib의 대화형 버전. 다양한 기능을 제공하고 디자인이 좋은 플롯 생성. Seaborn과 비교해 문법이 복잡함 기반

#### 8. 시각화 단계

- 1단계 : 시각화 라이브러리 임포트
- 2단계 : x축, y축에 표시할 데이터 정하기
- 3단계: 시각화를 위한 함수(예. plot())에 데이터 입력하기
- 4단계 : 그래프 보여 주기