

# 빅데이터 시각화 | 데이터와 시각화 이슈

“시각화란, 숫자를 공간에 배치해 보여줌으로써 그 패턴을 인지하게 만드는 것이다. 인간에겐 탁월한 패턴 인식 능력이 있다. 데이터 시각화는 통계 분석 기법으로는 도저히 알 수 없는 데이터의 이야기를 끌어낼 것이다” (참고: 비주얼라이즈 디스)

데이터를 기반으로 객관적 표현에 초점

## 시각화의 효과

- 자료로부터 정보를 습득하는 시간 절감으로 즉각적인 상황 판단이 가능
- 자료를 습득하는 사람들의 흥미를 유발하고 정보의 빠른 확산을 촉진
- 자료를 기억하는 데에도 도움

## 빅데이터 시각화 방법

| 시간시각화   | 분포시각화   | 관계시각화   | 비교시각화   | 공간시각화   |
|---|---|---|---|---|
| <ul style="list-style-type: none"><li>▪ 막대 그래프</li><li>▪ 누적 막대그래프</li><li>▪ 점 그래프</li></ul> | <ul style="list-style-type: none"><li>▪ 파이차트</li><li>▪ 도넛차트</li><li>▪ 트리맵</li></ul> <p>(Tree구조 분석)</p> <ul style="list-style-type: none"><li>▪ 누적연속 그래프</li></ul> | <ul style="list-style-type: none"><li>▪ 산점도(scatter plot)</li><li>▪ 버블 차트</li><li>▪ 히스토그램</li></ul> | <ul style="list-style-type: none"><li>▪ 히트맵</li><li>▪ 막대그래프</li><li>▪ 스타차트</li><li>▪ 평행 좌표계</li></ul> | <ul style="list-style-type: none"><li>▪ 지도 매핑</li></ul> |

## 시간시각화

시간에 걸쳐 진행되는 변화 또는 트렌드 추적에 사용

-분절형 : 데이터의 특정 시점 또는 특정 시간 구간 값(예-연도별 현황)나

-연속형 : 기온 변화와 같이 시간 간격이 짧은 데이터

시간 시각화 기법 : 막대 그래프, 누적 막대그래프, 점 그래프

### [ 막대 그래프 ]

[특징]

- 분절형 시간 시각화
  - 값들이 뚜렷한 차이를 보이는 경우
  - 수치를 길이로 표현해 절대값을 갖는 막대를 배치
- 시간에 따른 트렌드와 경향성을 확인

[해석]

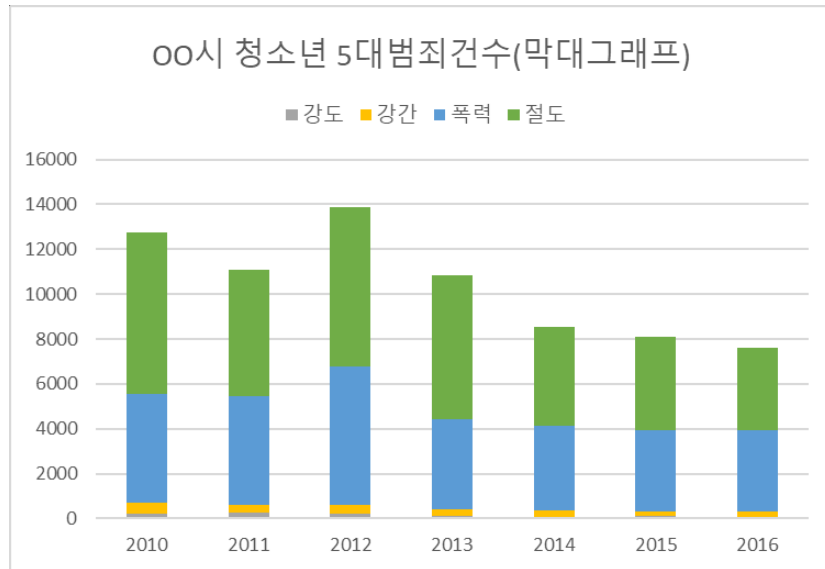
- 시간축(가로): 시간 순서대로 정렬된 시간의 특정 시점을 나타냄
- 값축(세로): 그래프의 크기, 범위를 나타냄

[유의점]

- 막대 값들의 차이가 미미 하거나 표시할 값(막대) 의 수가 많은 경우, 비교가 쉽지 않음
- 시각적 차이를 강조하기 위해 막대에 다양한 색상 적용
- 색상은 특정 상태나 범위를 나타냄



## [ 누적 막대그래프 ]



### [특징]

- 일반적인 막대그래프와 거의 비슷
- 한 구간에 해당하는 막대가 누적
- 한 구간이 몇 개의 세부 항목으로 나뉘면서도 전체의 합이 의미가 있을 때 누적 막대그래프 사용 → 한 구간의 각 세부항목은 질감/색상으로 구분

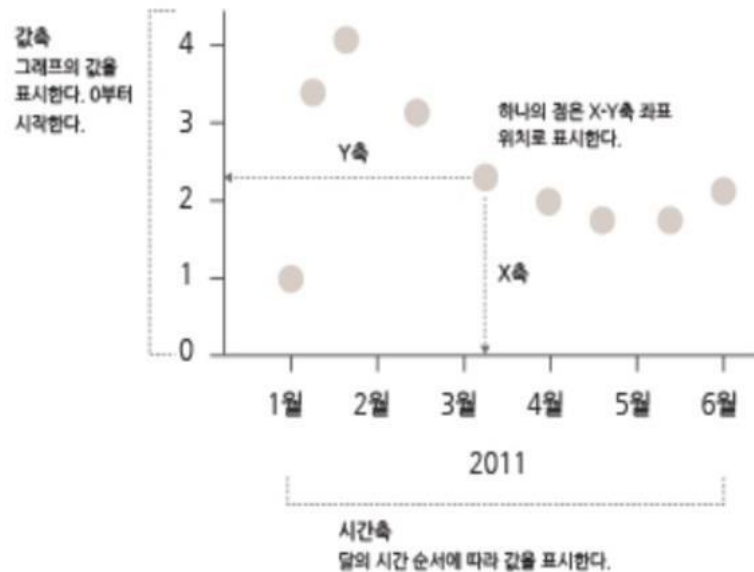
## [ 점 그래프 ]

### [특징]

- 면적을 표시할 필요가 없어서 더 적은 공간에 표현 가능
- 한 점에서 다음 점으로 변하는 점의 집중 정도와 배치에 따라 흐름을 파악하기에 용이
- 두 변수의 연관 관계를 보여줄 때 사용

### [해석]

- 시간축(가로) : 시간 순서대로 정렬된 시간의 특정 시점을 나타냄
- 값축(세로) : 그래프의 값을 표시
- 점 : 하나의 점은 X-Y축 좌표의 위치로 표시



## [ 연속형 시간 데이터 표현을 위한 연결된 점.선 그래프 ]

### [특징]

- 점그래프와 거의 같으나, 점 사이를 선으로 잇는다는 것이 차이점
- 선으로 표현되는 연속적인 데이터가 끊임없이 변화하는 현상의 추이를 확인
- 변수의 변화, 트렌드, 변화율 정보가 중요한 경우 사용

### [해석]

- 점 : 하나의 점은 X-Y축 좌표의 위치로 표시
- 선분 : 노드를 연결하는 선으로 노드 간의 변화 추이를 보여줌

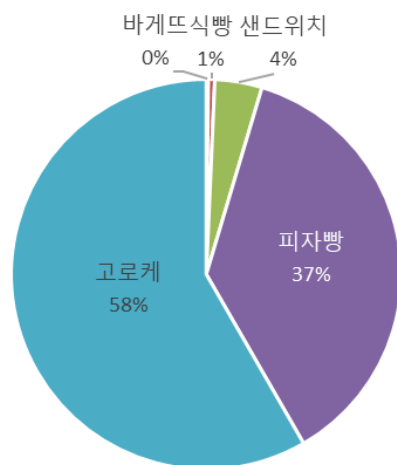


## 분포 시각화

전체의 관점에서 부분 간의 관계 분석

부분을 전부 합치면 1 또는 100%가 됨

비오는날 제빵류 판매비율



### [ 원(Pie) 차트 ]

[특징]

- 부분과 부분 간의 비율을 알아보는 데 사용
- 분포의 정도를 총합 100%로 나타내서 부분 간의 관계를 보여줌
- 육안으로 면적 가늠 후 시각(Visual angles) 비교 → 어떤 경우는 그래프 해석이 쉽지 않음

### 분포 시각화 기법

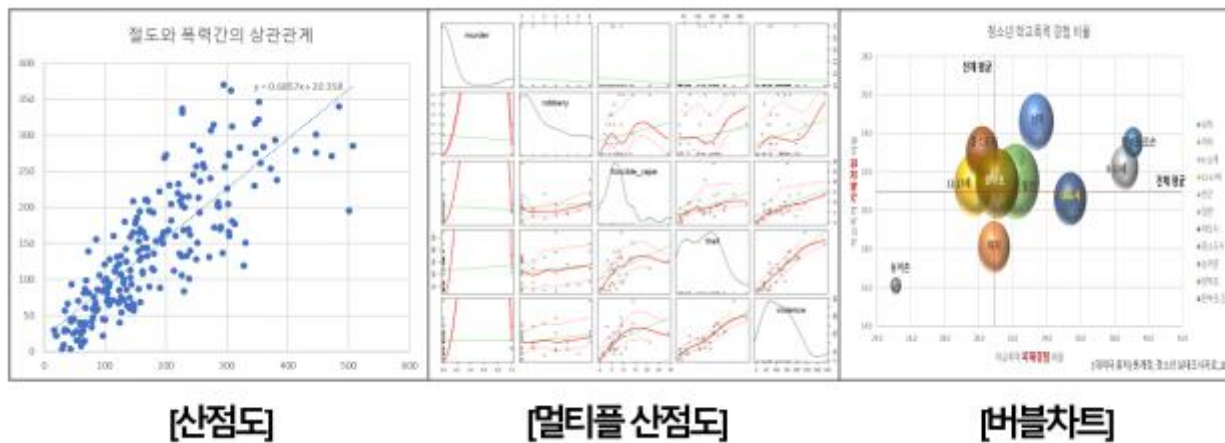
- 파이차트
- 도넛차트
- 트리맵 (Tree구조 분석)
- 누적연속 그래프

## 관계 시각화

데이터 간의 상관 관계 분석

한 수치의 변화를 모니터링하여 다른 수치의 변화 예측에 이용

### 관계 시각화 기법



## 비교 시각화

### 비교 시각화 필요성

- 여러 개의 변수를 다뤄야할 때 마주하는 첫번째 난관 → 시작점을 찾는 것
- 많은 데이터를 끊임 없이 생각하다 보면 너무 많은 변수와 세부 분류에 압도되어 분석이 어려워짐

### 비교 시각화

데이터를 전체적으로 한 눈에 볼 수 있도록 값을 색으로 나타내는 방법

### 효과

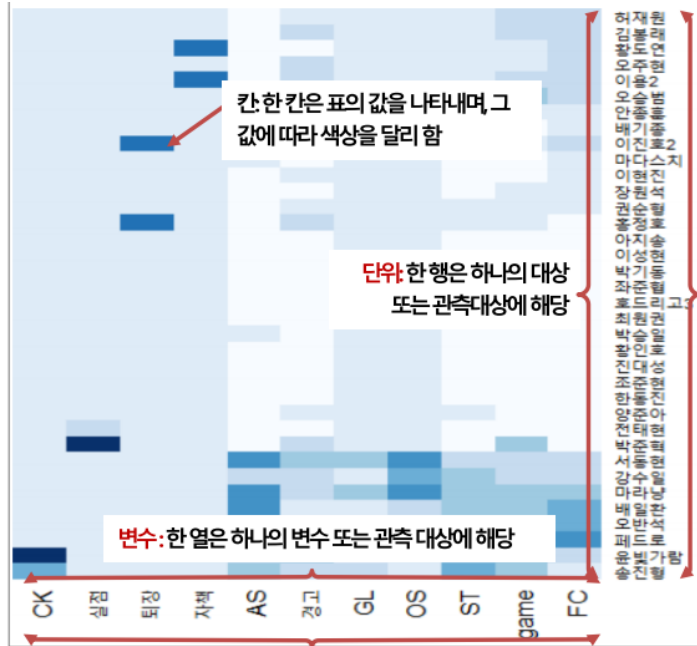
- 모든 데이터를 한번에 훑어본 다음 흥미로운 점을 짚고  
다른 점을 찾아가는 방향을 제시해 줄 수 있음
- 한 수치의 변화를 모니터링하여 다른 수치의 변화 예측에 이용

### 비교 시각화 기법

히트맵, 체르노프 페이스, 스타차트, 평행 좌표계, 다차원 척도법



## 히트맵(Heatmap)



### [특징]

- 시각화 기법에서 가장 많이 유용하게 쓰이는 그래프
- 여러가지 변수를 비교 가능
- 한 칸의 색상으로 데이터 값을 표현

### [해석]

- 하나의 대상에 해당하는 한 행을 왼쪽→오른쪽으로 보면서 모든 변수를 파악 가능
- 변수에 대응하는 한 열을 위→아래로 읽을 수 있음

### [사용시 유의점]

- 데이터가 지나치게 많을 경우 혼란스러울 수 있음
- 적당한 색상 선택, 약간의 정렬과정을 거쳐야 함

### [예시 데이터]

- 한국프로축구연맹, 제주 유나이티드 축구단 선수들의 K리그 개인기록, 2013