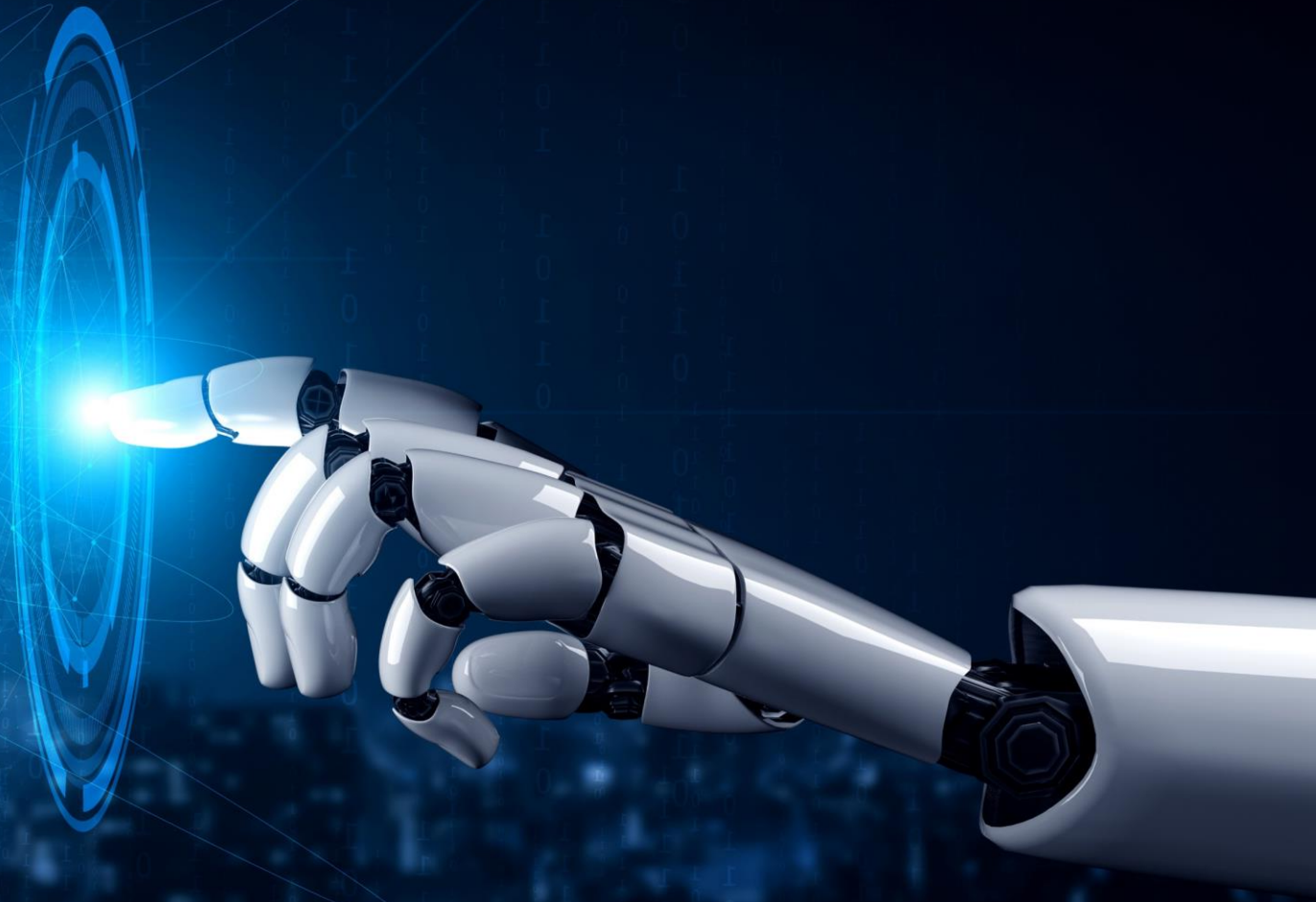


# 빅데이터 분석 시각화 개념

충북대학교 소프트웨어학과  
류관희



# 목 차

## ❖ Part 1. 빅데이터 시각화 정의

- 빅데이터 정의
- 빅데이터 경향
- 빅데이터 시각화 역사
- 빅데이터 시각화 과정

## ❖ Part 2. 빅데이터 시각화 필요성

- 빅데이터 시각화의 필요성

## ❖ Part 3. 빅데이터 시각화 종류

- 시간 시각화
- 분포 시각화
- 관계 시각화
- 공간 시각화



01

- 빅데이터 정의
- 빅데이터 경향
- 빅데이터 시각화 역사
- 빅데이터 시각화 과정

02

- 빅데이터 시각화의 필요성

03

- 시간 시각화
- 분포 시각화
- 관계 시각화
- 공간 시각화

# 빅데이터 정의(1)

## • 빅데이터(Big Data)란?

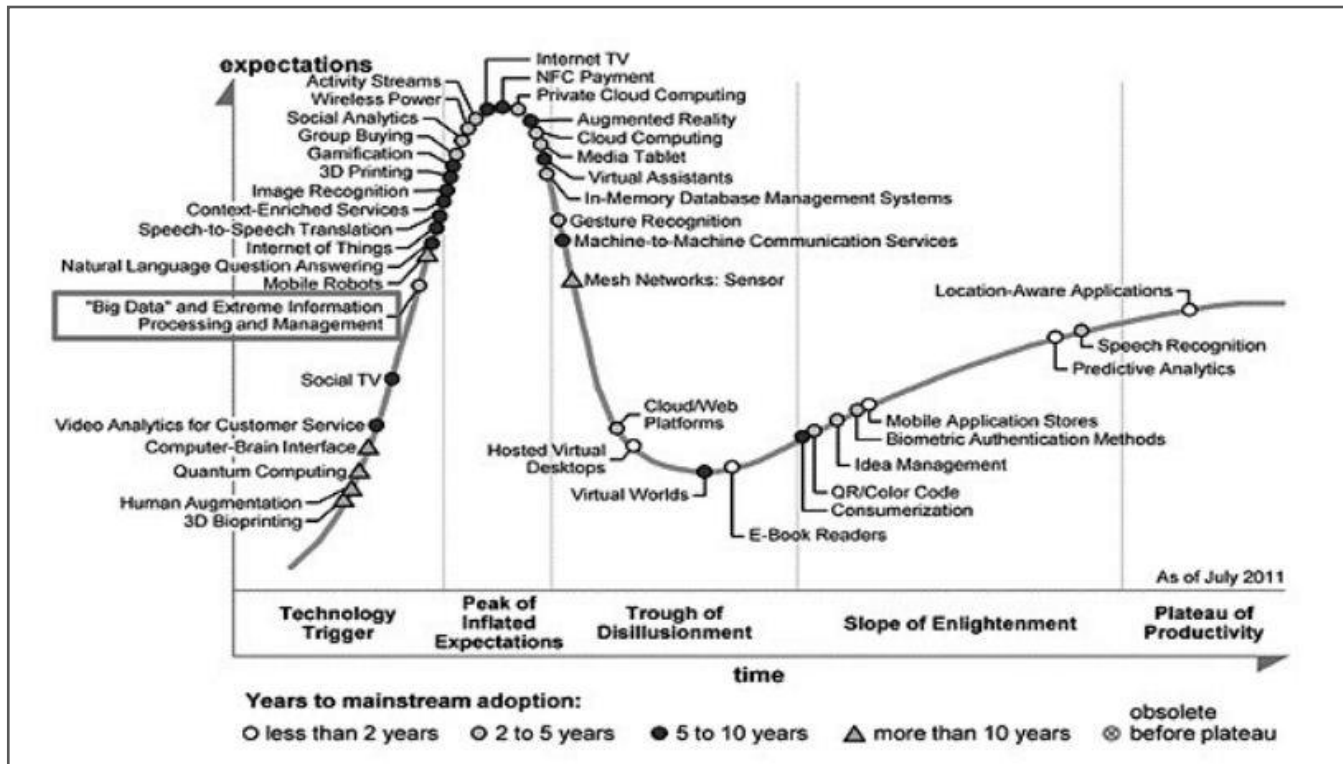
- 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합을 지칭 (SERI, 2010)

구 분	주 요 내 용
규모(Volume)의 증가	<ul style="list-style-type: none"> <li>• 기술적인 발전과 IT의 일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증, 제타바이트(ZB) 시대로 진입</li> </ul>
다양성(Variety) 증가	<ul style="list-style-type: none"> <li>• 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터 종류의 증가</li> <li>• 텍스트 이외의 멀티미디어 등 비정형화된 데이터 유형의 다양화</li> </ul>
속도(Velocity) 증가	<ul style="list-style-type: none"> <li>• 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간 정보 증가</li> <li>• 실시간성으로 인한 데이터 생성, 이동(유통) 속도의 증가</li> <li>• 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터 처리 및 분석 속도가 중요</li> </ul>



## 빅데이터 정의(2)

- 빅데이터(Big Data)  
2011년도 가트너의 어머징 기술 하이프 사이클(Hype Cycle)



# 빅데이터 경향(1)

## • 빅데이터(Big Data) 경향

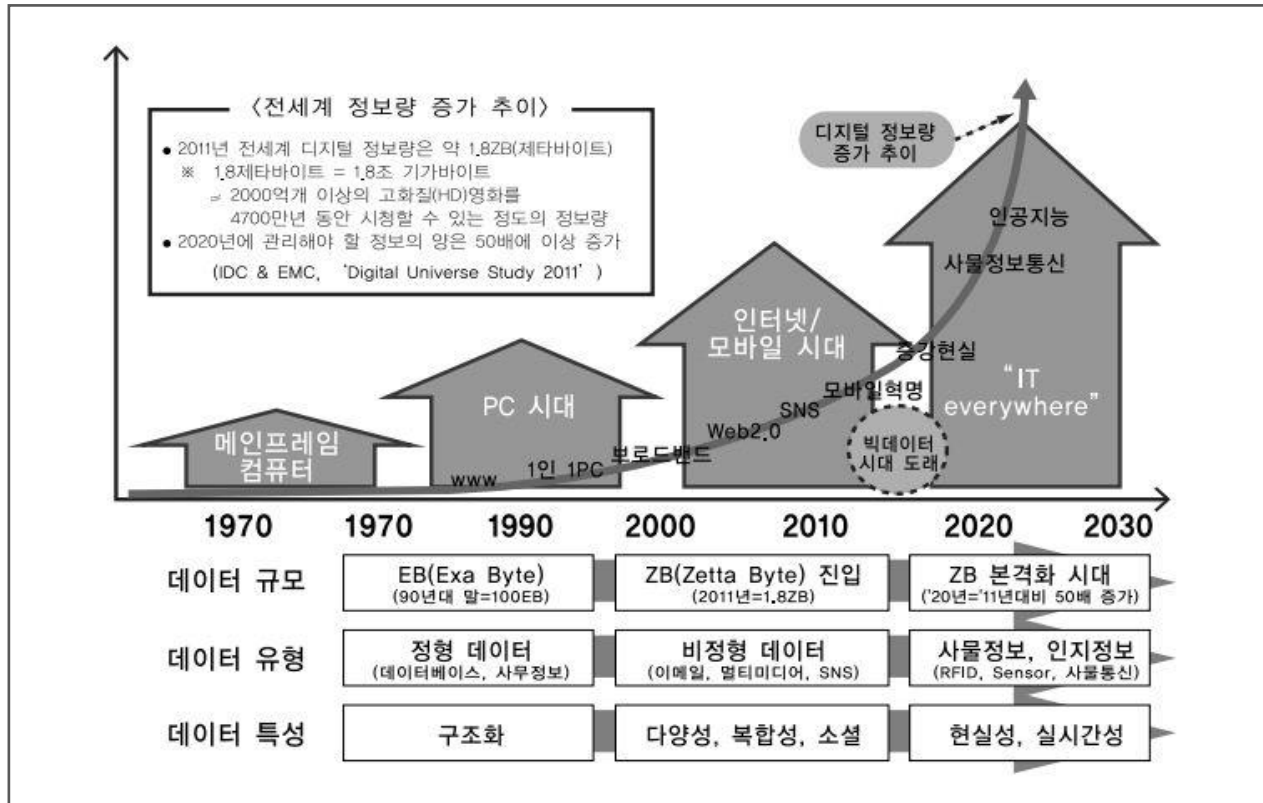
정보사회의 패러다임 변화와 힘의 이동

	PC시대	인터넷시대	모바일시대	스마트시대
패러다임 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물정보화
IT 이슈	PC, PC통신, 데이터베이스	초고속 인터넷, www, 웹서버	모바일 인터넷, 스마트폰	빅데이터, 차세대PC, 사물네트워크(M2M)
핵심분야 (서비스)	PC, OS	포털, 검색엔진, Web2.0	스마트폰, 앱서비스, SNS	미래전망, 상황인식, 개인맞춤형 서비스
대표 기업	MS, IBM 등	구글, 네이버, 유튜브 등	애플, 페이스북, 트위터 등	?
IT비전	1인 1PC	클릭 e-Korea	손안의 PC, 소통	IT everywhere, 新가치창출

## 빅데이터 경향(2)

### • 빅데이터(Big Data) 경향

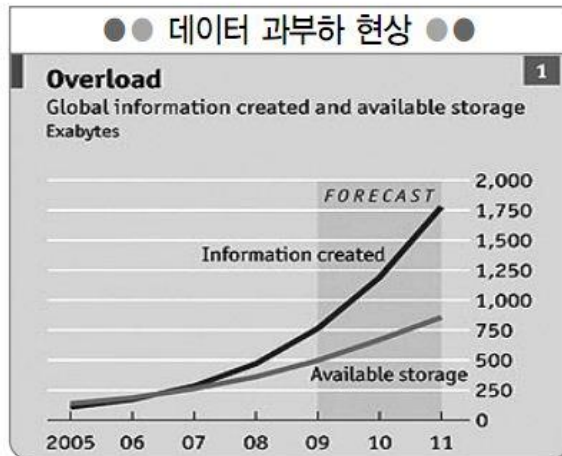
ICT(Information & Communication Technology)발전에 따른  
데이터의 변화와 방향



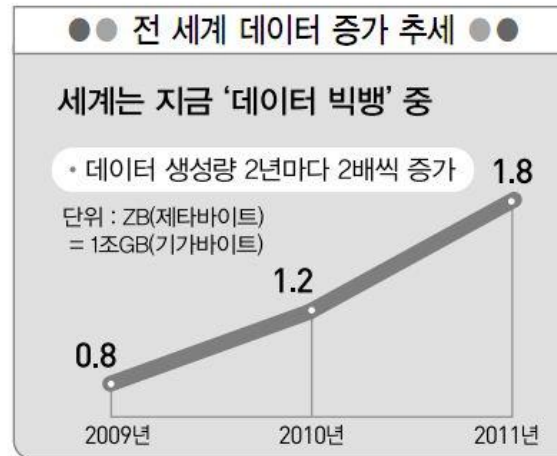
[2]김성태, "새로운 미래는 여는 빅데이터 시대", 한국정보화진흥원, 2013.

## 빅데이터 경향(3)

- 빅데이터(Big Data) 경향
  - 제타바이트 시대
    - 인터넷이 일상화된 최근 10년 사이, 인류는 디지털 데이터가 증폭하는 데이터 홍수 현상에 직면
    - 디지털 정보량의 기하급수적인 증가에 따라 대규모 데이터가 종대 이슈로 부각하며 **‘빅데이터’**라는 용어가 등장



※ 출처 : IDC, Economist(2010) 재인용



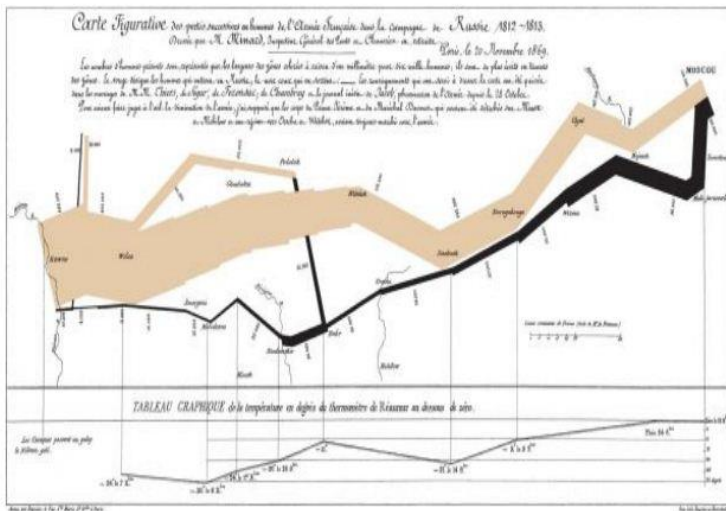
※ 그림출처 : 중앙일보(2011)



# 빅데이터의 시각화 역사(1)

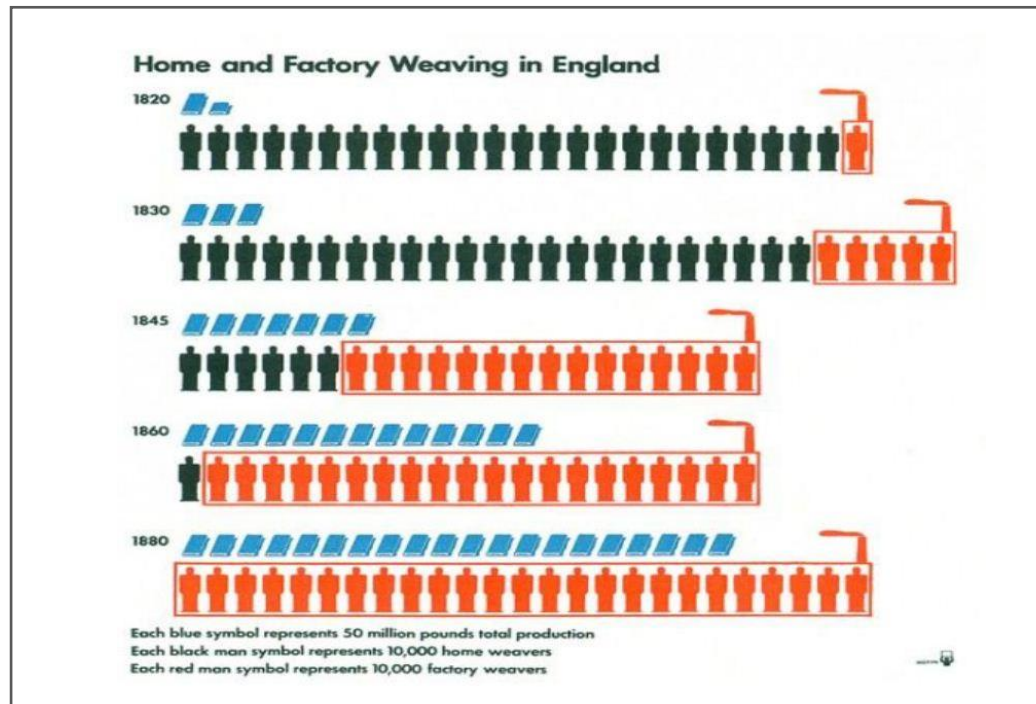
- 1800s : 커뮤니케이션 수단
  - Playfair, Nightingale, Snow, Minard

Minard : Napoleon's Russian campaign in 1812/13



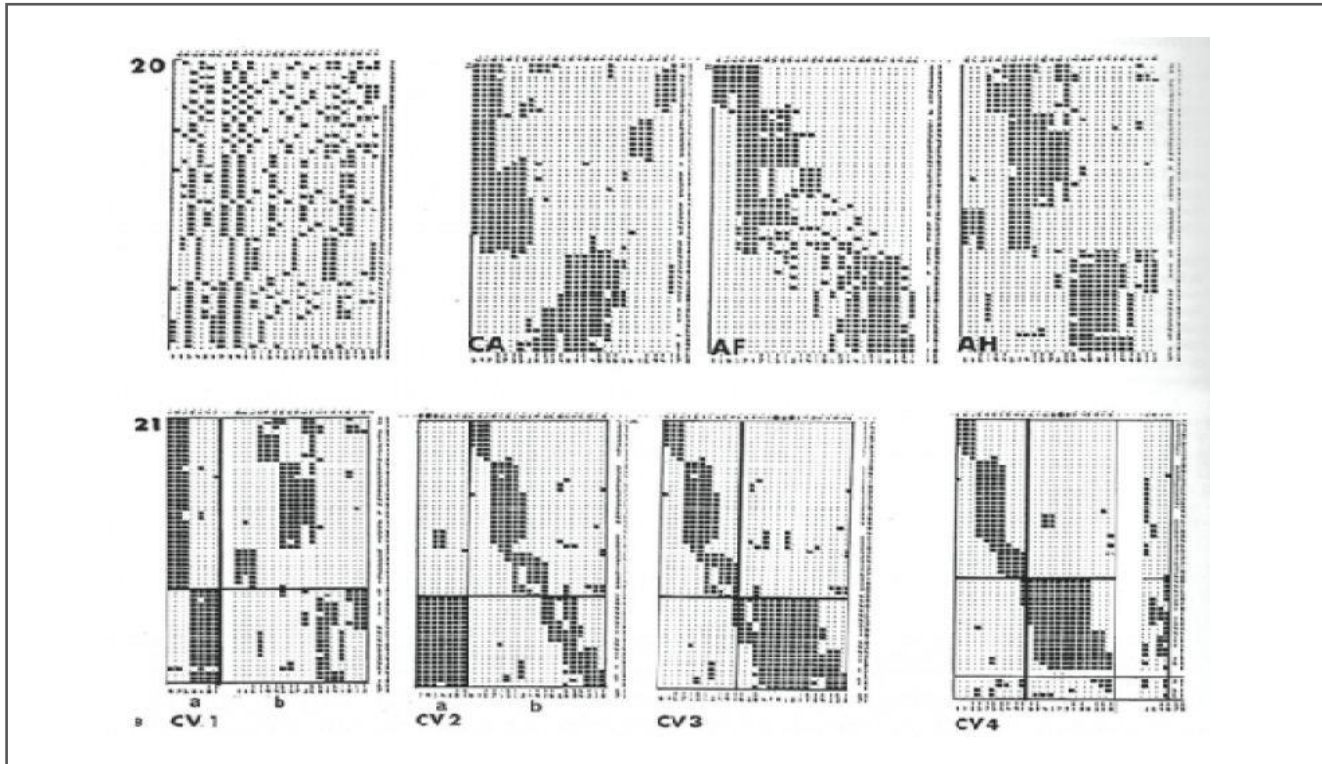
## 빅데이터의 시각화 역사(2)

- 1920s – 1930s : 읽기 쉬움&의미 전달  
(ISOTYPE(International System of Typographic Picture Education))
  - Neurath



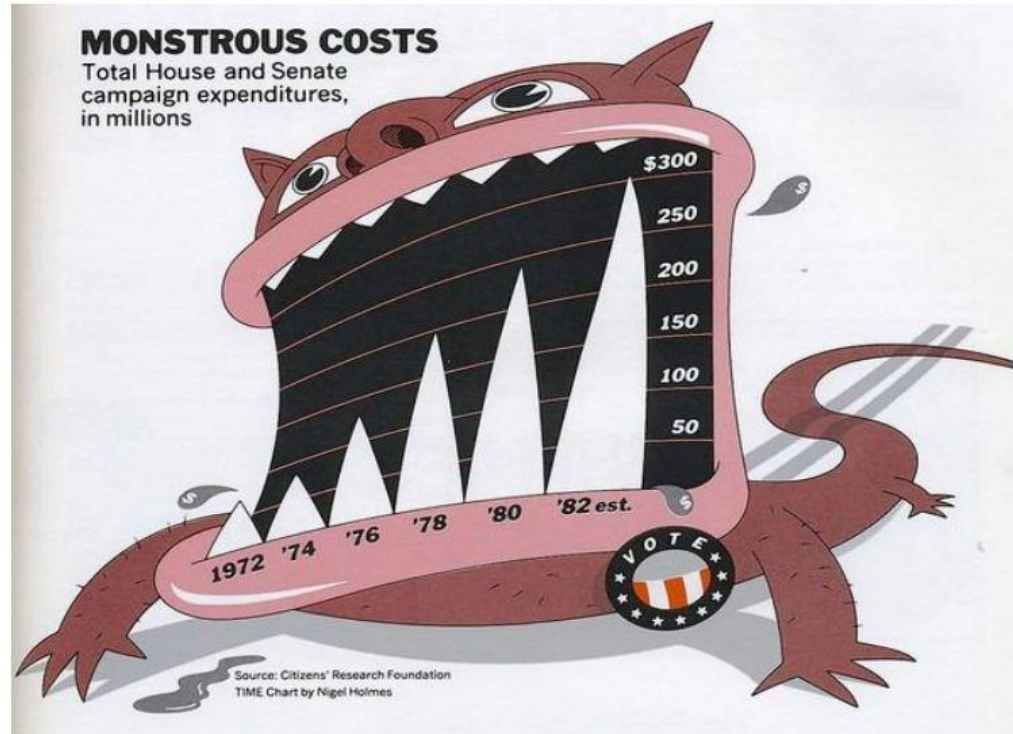
## 빅데이터의 시각화 역사(3)

- 1960s – 1970s : 분석
  - Bertin and Tukey



## 빅데이터의 시각화 역사(4)

- 1960s – 1970s : Information 그래픽
  - Holmes, Tufte

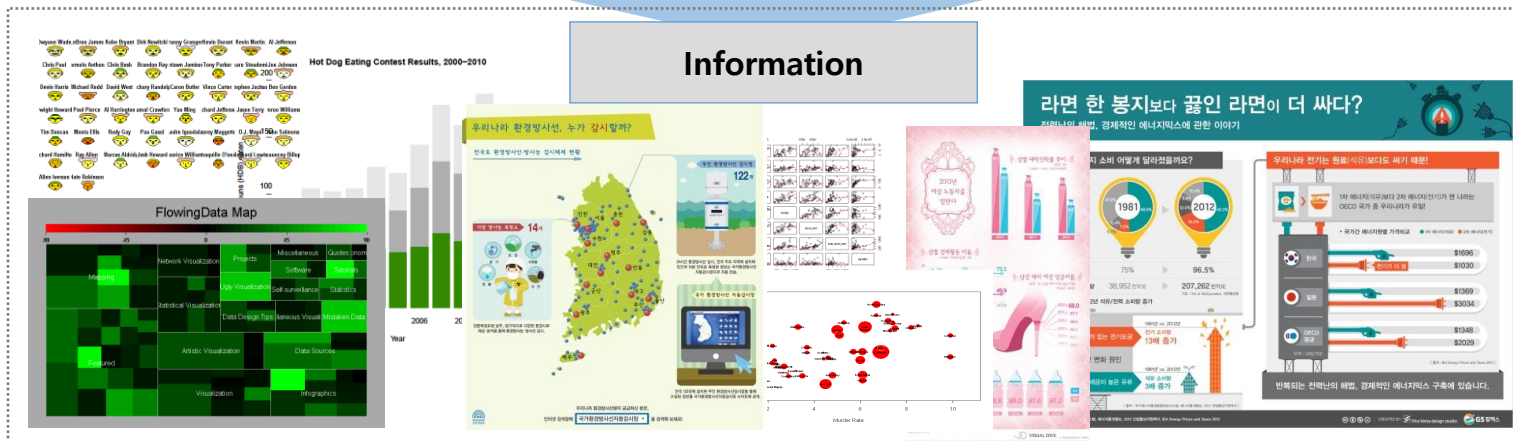
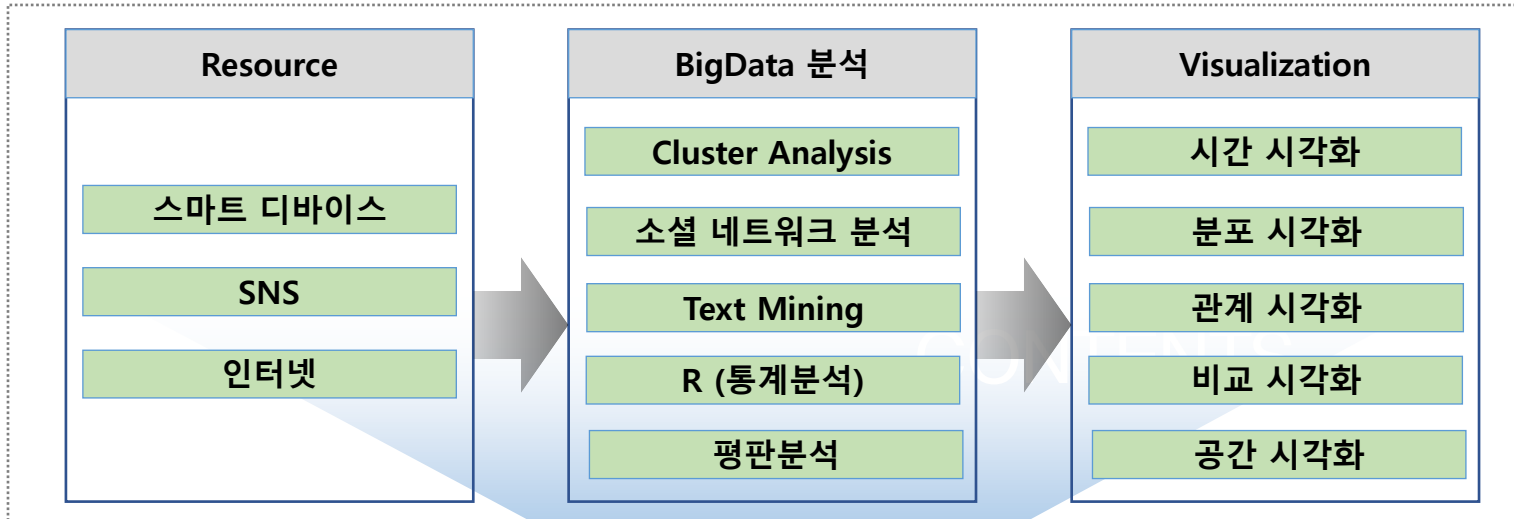




[illegible]



# 빅데이터 시각화 과정



## 문제풀이

- 빅데이터에 대해 정의 하시오.
- 빅데이터 경향에 대해 설명 하시오.
- 빅데이터 시각화 과정에 대해 설명 하시오.

## 요약

- 빅데이터는 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터 집합을 지칭 한다.
- 정보 사회의 패러다임 변화, ICT 발전에 따른 데이터 방향 변화, 데이터 홍수 현상의 제타바이트의 시대가 부각되며 빅데이터 시대를 정의 한다.
- 빅데이터 시각화 역사에 대해 알아보며 각 시대별 시각화 의미에 대해 설명한다.
- 빅데이터 시각화 과정은 데이터 수집, 데이터 분석, 적절한 시각화 기법을 이용한 데이터 시각화 방법으로 진행된다.

01

- 빅데이터 정의
- 빅데이터 경향
- 빅데이터 시각화 역사
- 빅데이터 시각화 과정

02

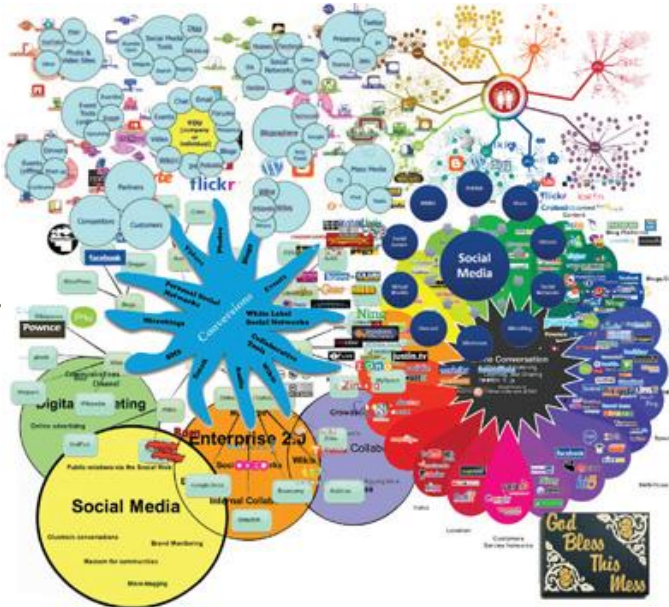
- **빅데이터 시각화  
의 필요성**

03

- 시간 시각화
- 분포 시각화
- 관계 시각화
- 공간 시각화

# 빅데이터 시각화 필요성

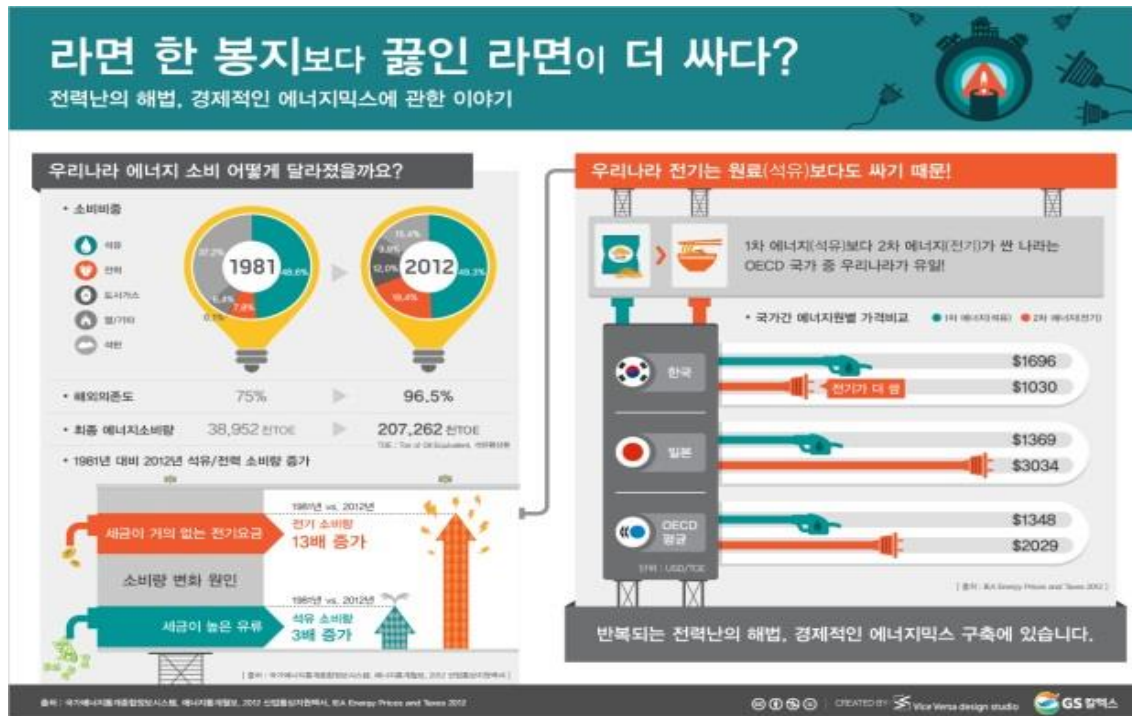
- 빅데이터 시각화 필요성
  - 많은 양의 빅데이터를 시각화 하지 않으면 패턴을 볼 수 없음





# 빅데이터 시각화 필요성

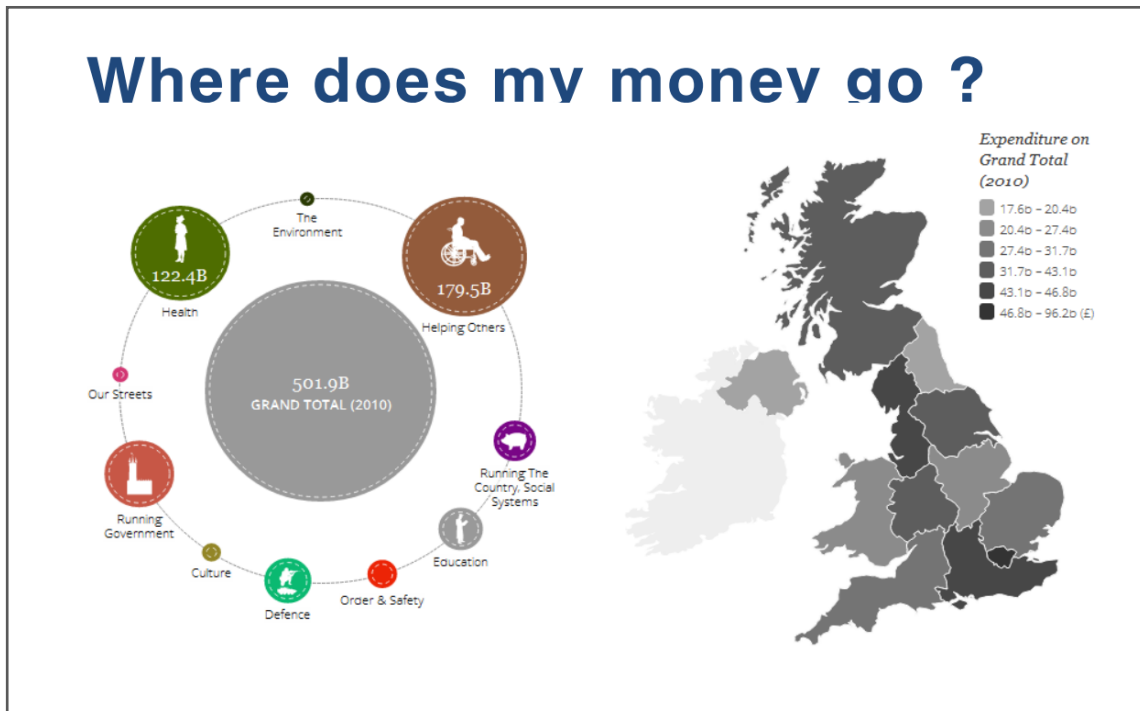
- 빅데이터 시각화 필요성
  - 빅데이터를 시각화 하면 하나의 화면에 필요한 데이터 요소를 모두 적합할 수 있음



<http://info-graphics.kr/>

# 빅데이터 시각화 필요성

- 빅데이터 시각화 필요성
  - 빅데이터를 시각화 하면 한 화면에 깊고 폭 넓은 데이터 세트를 효과적으로 표시 할 수 있음



## 문제풀이

- 빅데이터 시각화의 필요성에 대해 설명하시오.

## 요약

- 빅데이터를 시각화 하지 않으면 데이터 속의 숨은 패턴을 볼 수 없어 데이터를 분석하기에 어려움이 있다.
- 빅데이터는 양 뿐만 아니라 다양한 자료를 포함하고 있어 다양한 분석 결과들을 한 화면에 모두 담기 위해 빅데이터 시각화 기술의 발전이 수반되어야 한다.
- 빅데이터를 분석한 결과는 시각화 하면 보다 효과적으로 데이터를 전달 할 수 있다.

01

- 빅데이터 정의
- 빅데이터 경향
- 빅데이터 시각화 역사
- 빅데이터 시각화 과정

02

- 빅데이터 시각화의 필요성

03

- 시간 시각화
- 분포 시각화
- 관계 시각화
- 공간 시각화



## 시간 시각화(1)

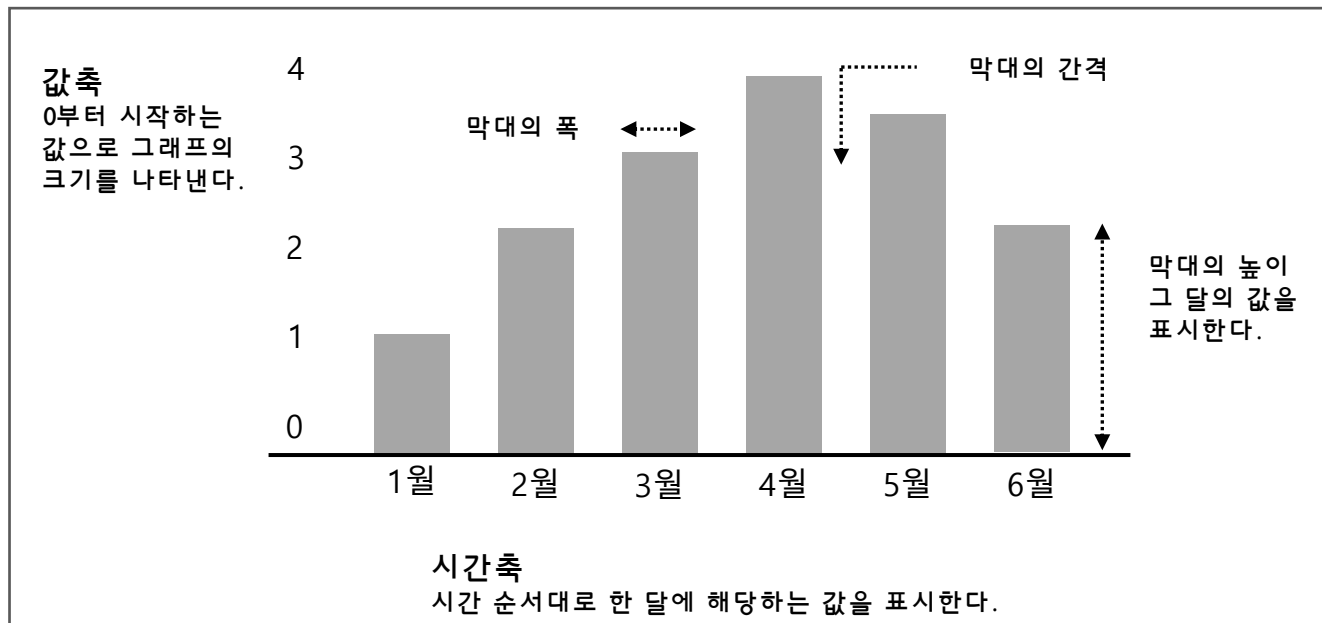
### • 시간 시각화란?

- 시간에 관련된 데이터를 시계열 데이터라고 하며 가장 큰 특징은 트렌드, 즉 경향성임
- 시간데이터:분절형, 연속형
  - 시간나눔:막대 그래프, 누적 막대 그래프, 점
  - 연속형 데이터:시계열 그래프, 계단식 그래프, LOESS 곡선 최적화

## 시간 시각화(2)

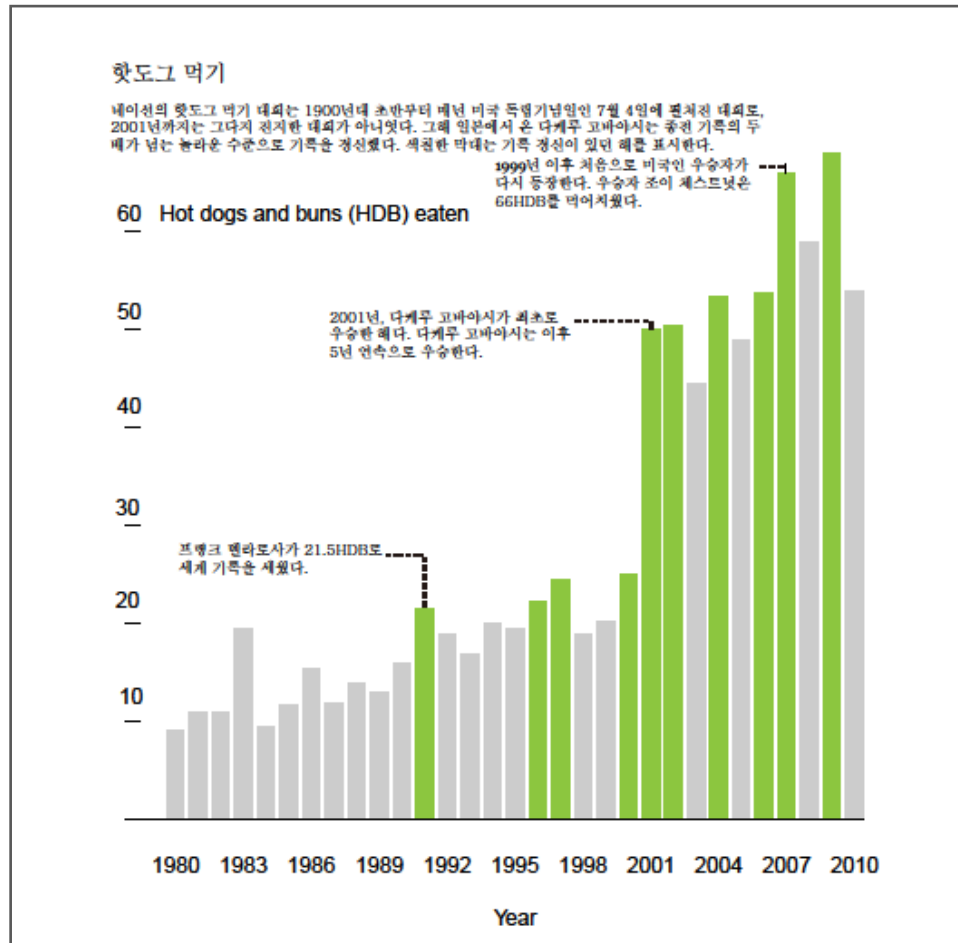
### • 막대 그래프: 시간축, 값 축

- 시간 축(x축) : 시간 순서대로 정렬된 시간의 특정 시점을 나타냄
- 값 축 (y축) : 그래프의 크기 범위를 나타냄



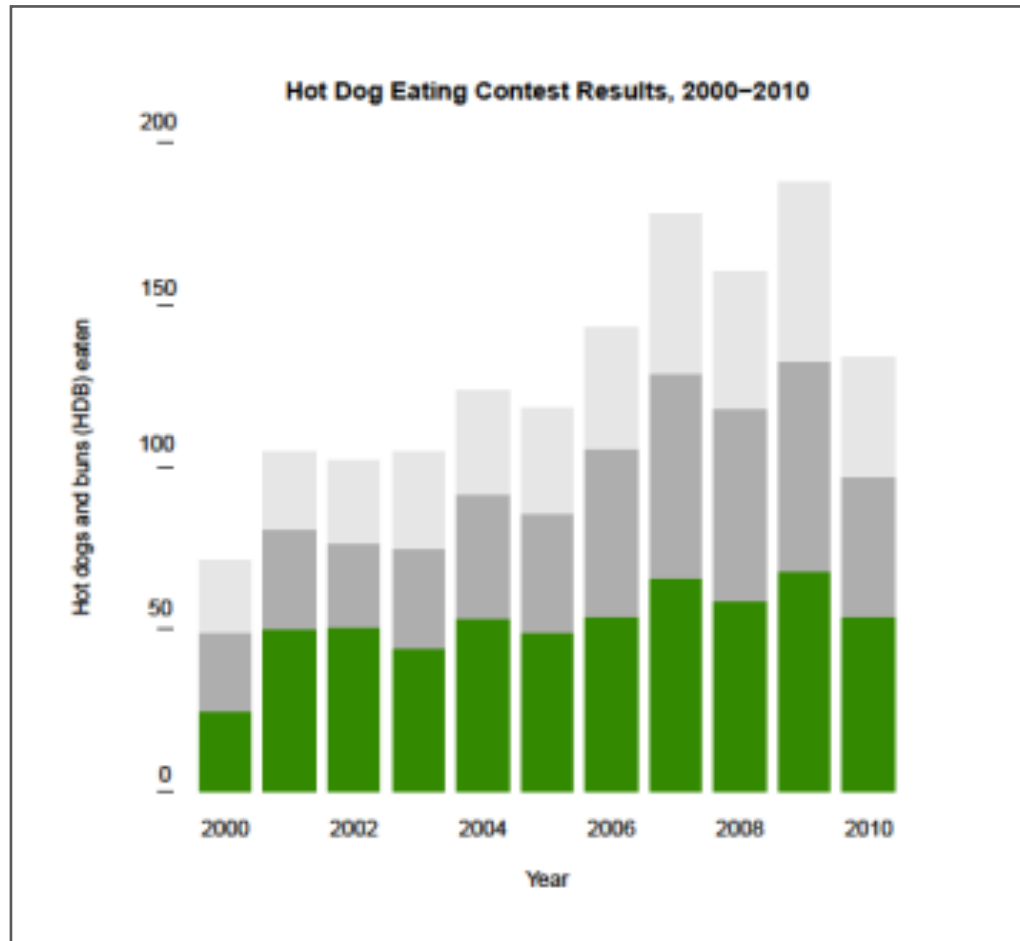
## 시간 시각화(3)

### • 일러스트로 보강



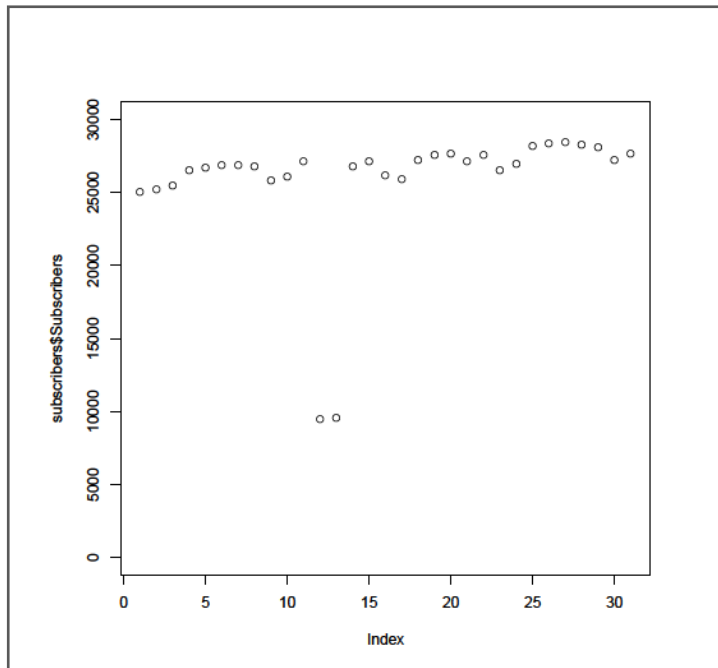
## 시간 시각화(4)

- 누적 막대 그래프 (일러스트레이트)

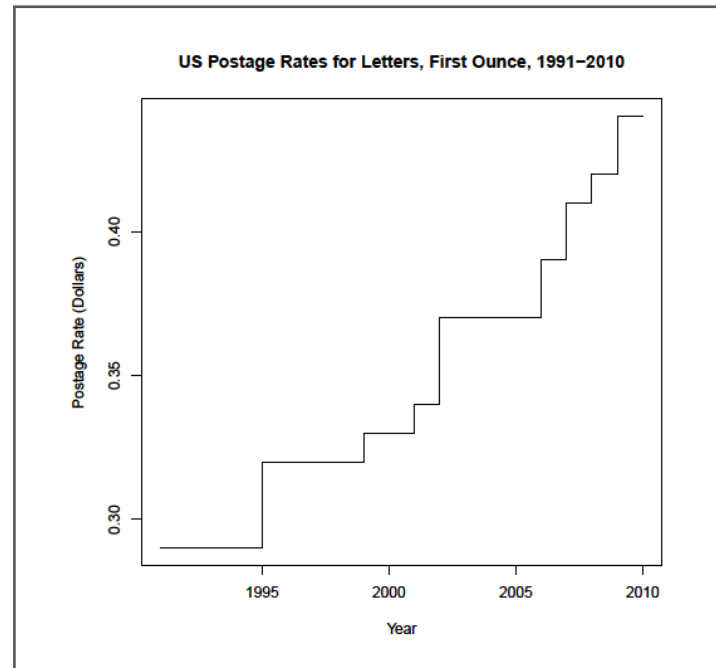


## 시간 시각화(5)

- 점 그래프(R)



- 계단식 그래프(R)





## 분포 시각화(1)

### • 분포 시각화란?

- 일반적인 특징: 최대, 최소, 전체분포
- 분포 데이터는 크게 전체의 부분, 시간에 따른 분포로 구분 함
  - 전체의 부분  
: 파이 차트, 도넛 차트, 누적 막대 그래프, 트리맵
  - 시간에 따른 분포  
: 누적 연속 그래프, 평행 좌표 그래프

## 분포 시각화(2)

### • 전체의 부분

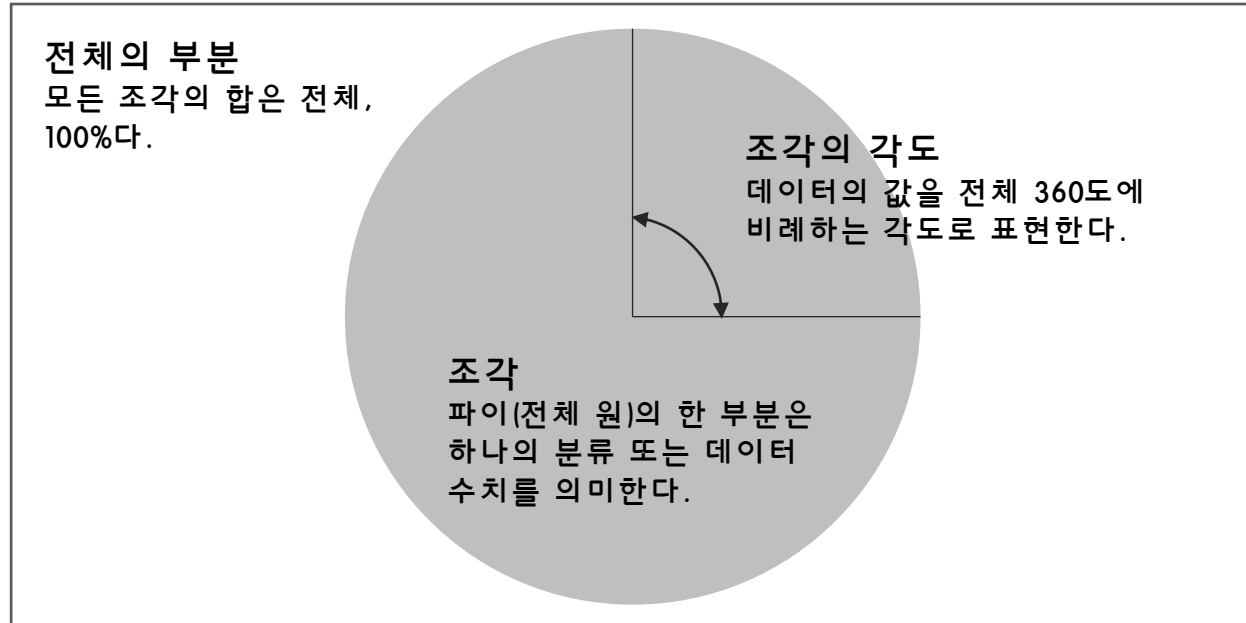
- 각 부분을 전부 합치면 1 또는 100%가 됨
- 이러한 데이터 특성에 맞게 전체의 관점에서 각 부분 간의 관계를 보여줘야 함

구 분	내 용
파이 차트	<ul style="list-style-type: none"> <li>• 원형의 각도를 기준으로 자른 조각 모형으로 표현</li> <li>• 하나의 조각은 전체의 한 부분을 나타냄</li> </ul>
도넛 차트	<ul style="list-style-type: none"> <li>• 파이 차트와 유사하지만 중심부를 잘라내어 도넛 모양으로 표현</li> <li>• 길이로 값의 차이를 나타냄</li> </ul>
누적 막대 그래프	<ul style="list-style-type: none"> <li>• 한 부분에 해당하는 막대를 누적으로 쌓아서 표현</li> </ul>
트리 맵	<ul style="list-style-type: none"> <li>• 영역 기반의 시각화</li> <li>• 각 사각형의 크기가 수치를 나타냄</li> </ul>

## 분포 시각화(3)

### • 전체의 부분-파이 차트

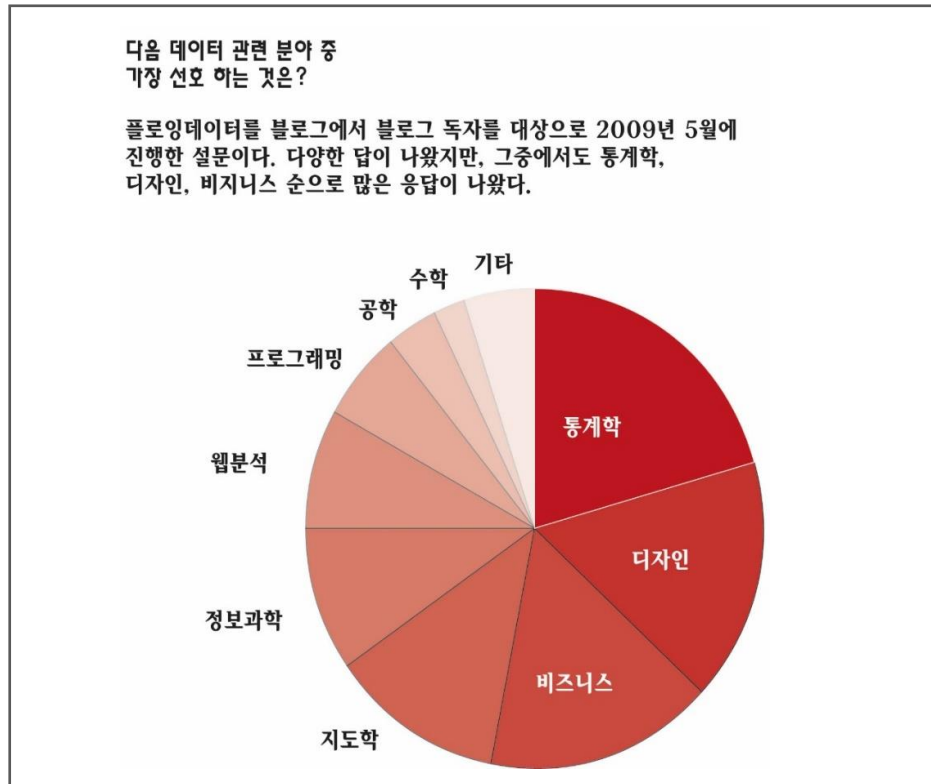
- 공식적인 최초의 파이 차트는 윌리엄 플레이 페어가 만듦
- 파이 차트는 원형을 기준으로 자른 조각으로 표현
- 하나의 조각은 전체의 한 부분을 나타냄



## 분포 시각화(3)

### • 전체의 부분-파이 차트

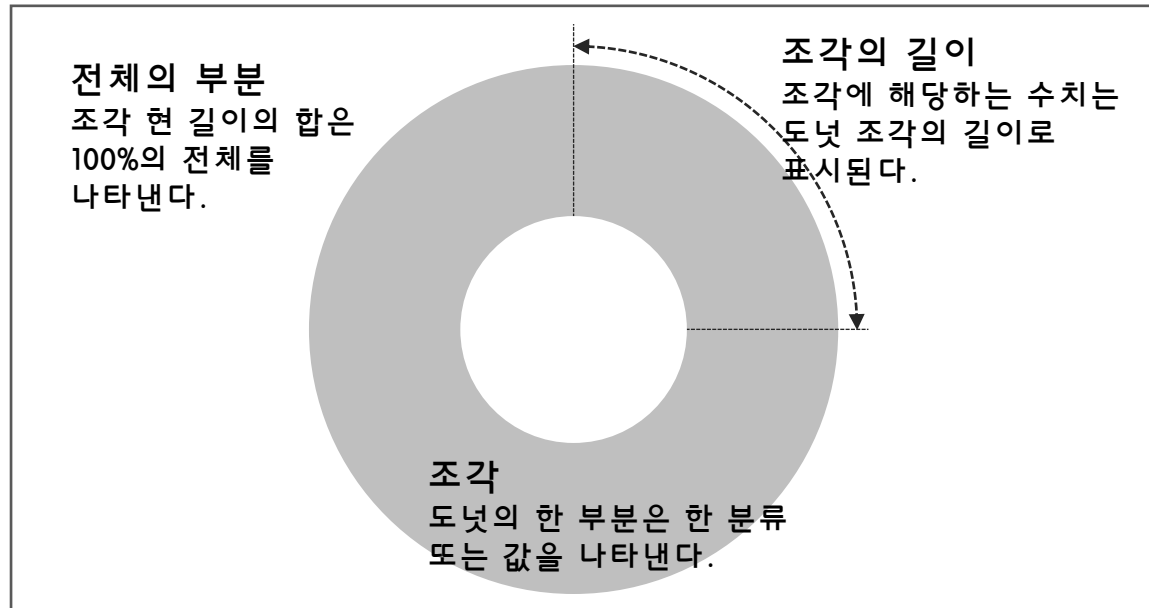
- 일러스트레이터를 위용한 파이 차트 생성



## 분포 시각화(4)

### • 전체의 부분-도넛 차트

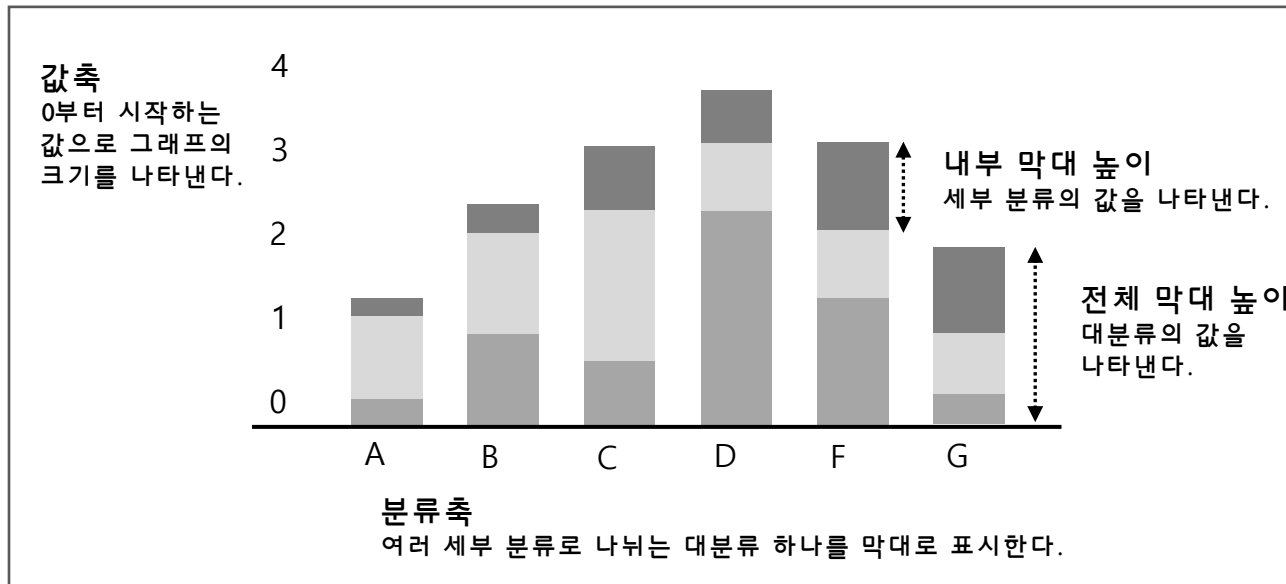
- 파이 차트와 마찬가지로 수치를 각도로 표시
- 중심부를 잘라 도넛 모양으로 표현
- 중심에 구멍이 뚫려 있기 때문에 각도보단 길이로 값의 차이를 인식



## 분포 시각화(5)

### • 전체의 부분-누적 막대 그래프

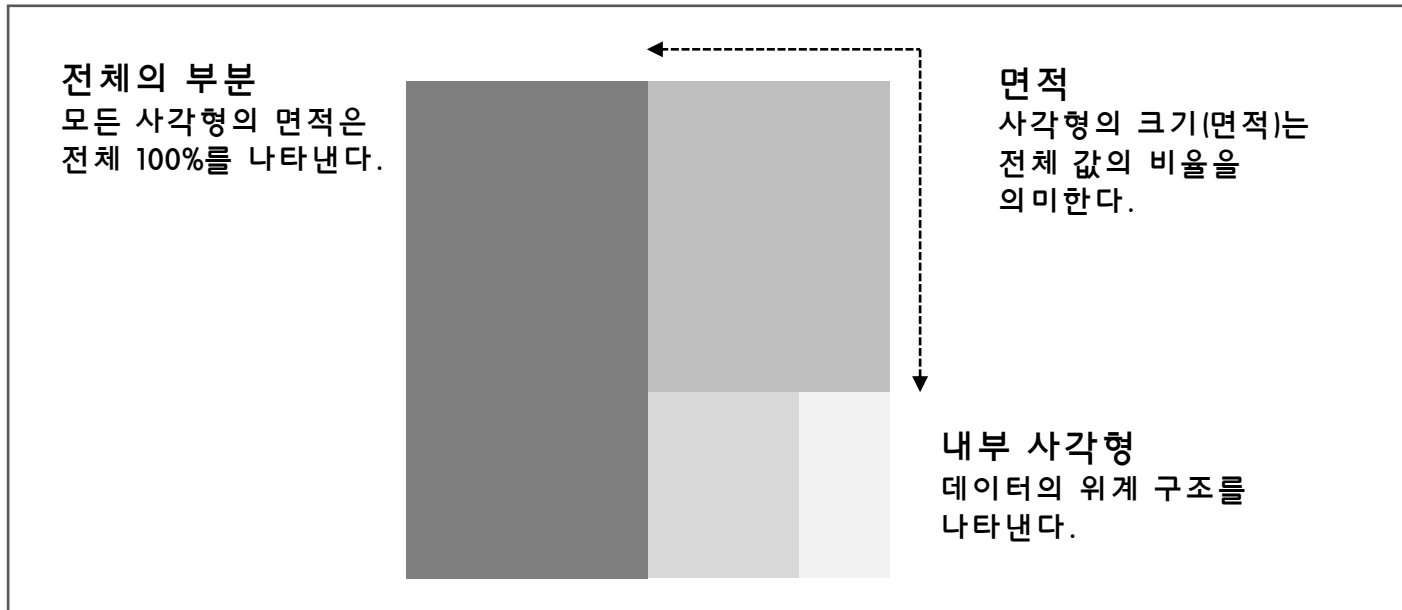
- 사물의 양을 막대 모양의 길이로 표시
- 데이터를 한눈에 알아보기 쉽고 이해하기 편함



## 분포 시각화(6)

### • 전체의 부분-도넛 차트

- 영역 기반의 시각화로, 각 사각형의 크기가 수치를 나타냄
- 단순 분류별 분포 시각화에도 쓸 수 있지만, 위계 구조가 있는 트리 구조에 데이터를 표시할 때 용이하게 활용





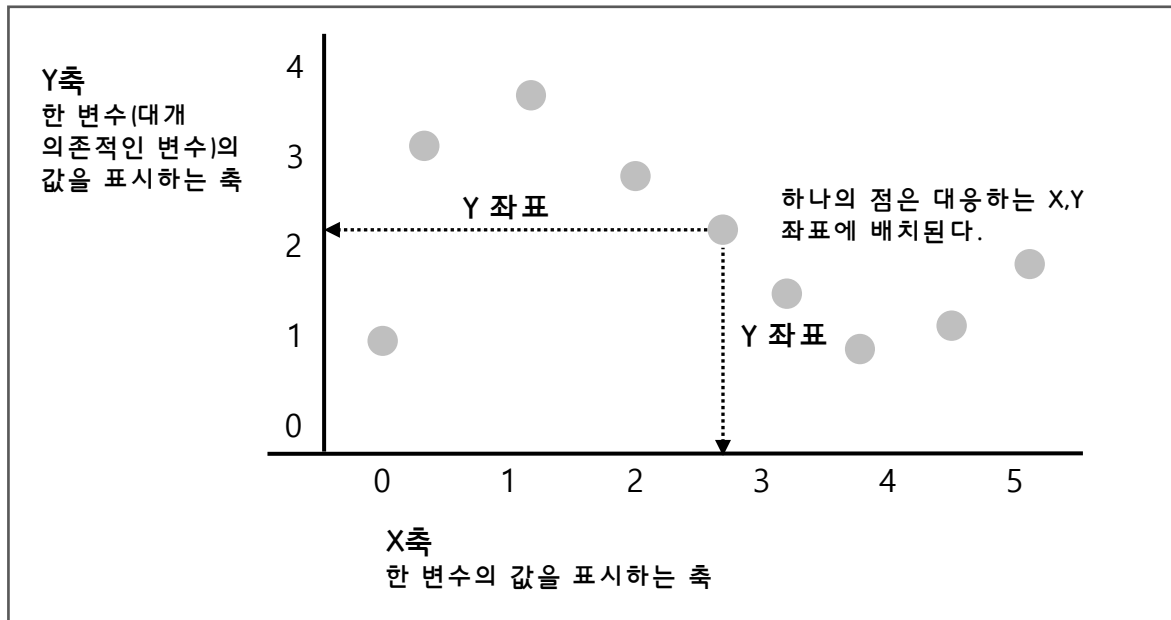
## 관계 시각화(1)

- 관계 시각화란?
  - 일반적인 특징 : 데이터의 관계, 분포, 비교
  - 관계 데이터는 크게 데이터의 관계를 설명하는 **상관관계**, 데이터의 분포를 나타내는 **분포**, 데이터를 비교하는 **비교**로 표현
    - 상관관계:캐스터 플롯,스캐터 플롯 행렬구조, 버블차트
    - 분포: 스템 플롯, 히스토그램, 연속 밀도 함수
    - 비교:히스토그램 행렬 구조, 스몰 멀티플

## 관계 시각화(2)

### • 상관관계-스캐터 플롯

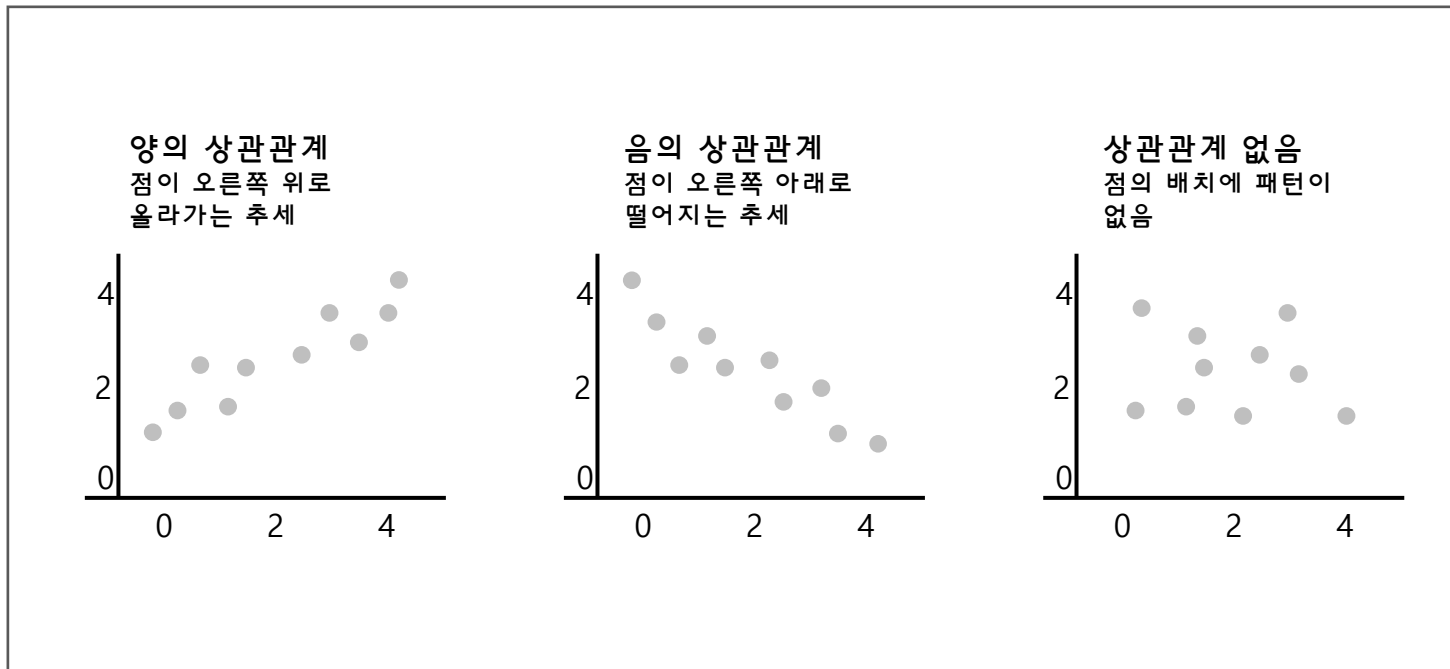
- 스캐터 플롯: X축, Y축
- 두 개 변수의 상관관계를 보여줌



## 관계 시각화(2)

### • 상관관계-스캐터 플롯

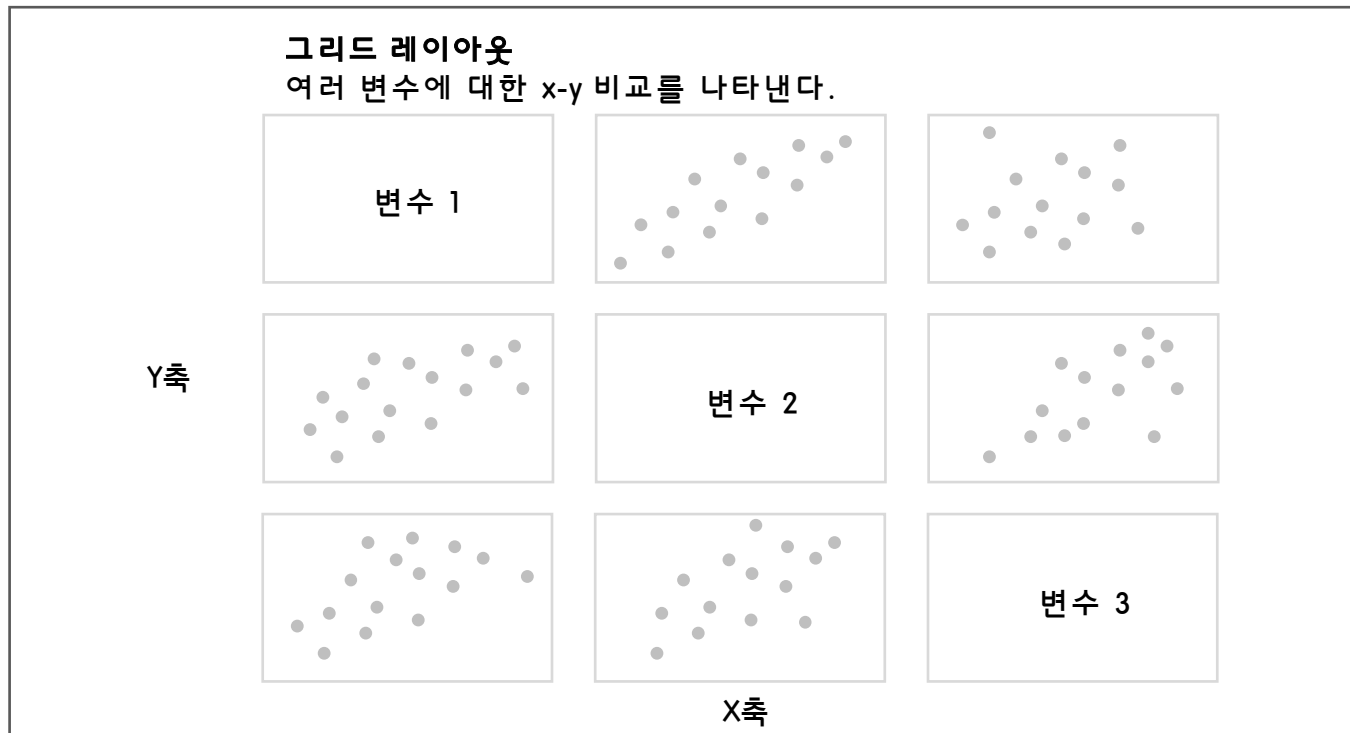
- 스캐터 플롯: X축, Y축
- 두 개 변수의 상관관계를 보여줌



## 관계 시각화(3)

### • 상관관계-스캐터 플롯 행렬

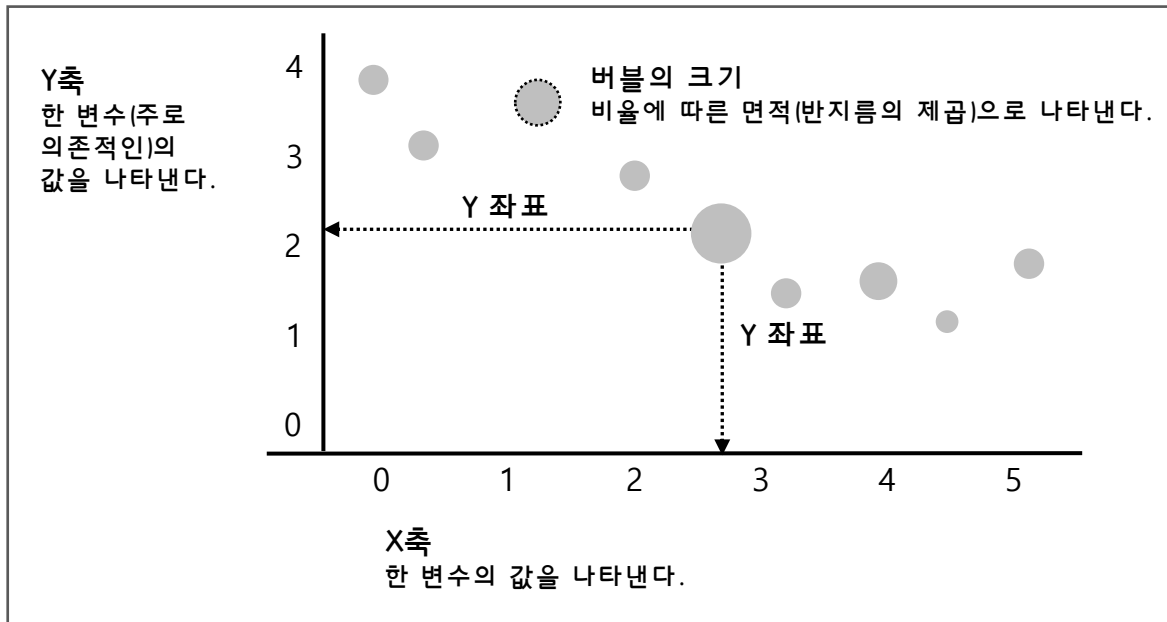
- 여러 변수의 관계를 보여줌



## 관계 시각화(4)

### • 상관관계-버블 차트

- X축, Y축, 버블의 크기
- 스캐터 플롯에 버블의 크기로 세 번째 변수(차원)을 나타내는 그래프



This scatter plot illustrates the relationship between the Murder Rate (X-axis) and the Burglary Rate (Y-axis) across various US states. The X-axis ranges from 0 to 10, and the Y-axis ranges from 200 to 1200. The size of each bubble represents the population of the state. The plot shows a positive correlation, where states with higher murder rates generally have higher burglary rates. States like North Carolina, Texas, and California are among the highest in both rates, while states like North Dakota and South Dakota are among the lowest.

State	Murder Rate (approx.)	Burglary Rate (approx.)
North Carolina	6.8	1180
Texas	6.2	950
California	6.5	700
Florida	5.0	900
Illinois	6.0	600
New York	4.5	350
North Dakota	1.5	300
South Dakota	2.5	300

## 관계 시각화(5)

### • 분포-스텝 플롯

- 단위 숫자는 왼쪽에 이어지는 숫자를 오른쪽에 기입
- 데이터의 분포를 확인하기 위한 그래프로 줄기와 잎 그림 이라고 함

세계은행에서 발표한 2008년  
세계 출생률의 스텝 플롯

The decimal point is at the |

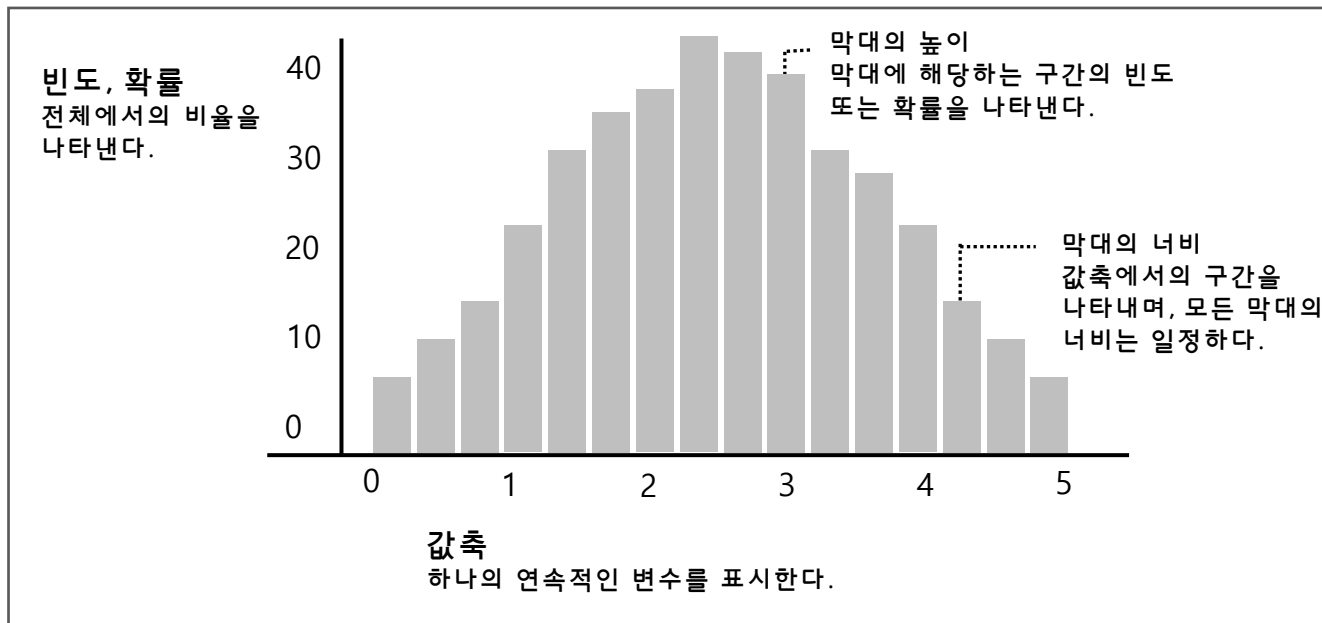
8		2371334468999
10		01223455566999001222334555777889
12		000111111356789993789
14		0034566788991237
16		227779123677889
18		00233677888900448
20		0024445688912455679
22		0057834579
24		11456677771347
26		31335667
28		014999
30		124234
32		1449069
34		556049
36		8890
38		023455823468
40		23125
42		699
44		17
46		252
48		
50		
52		5



## 관계 시각화(6)

### • 분포-히스토그램

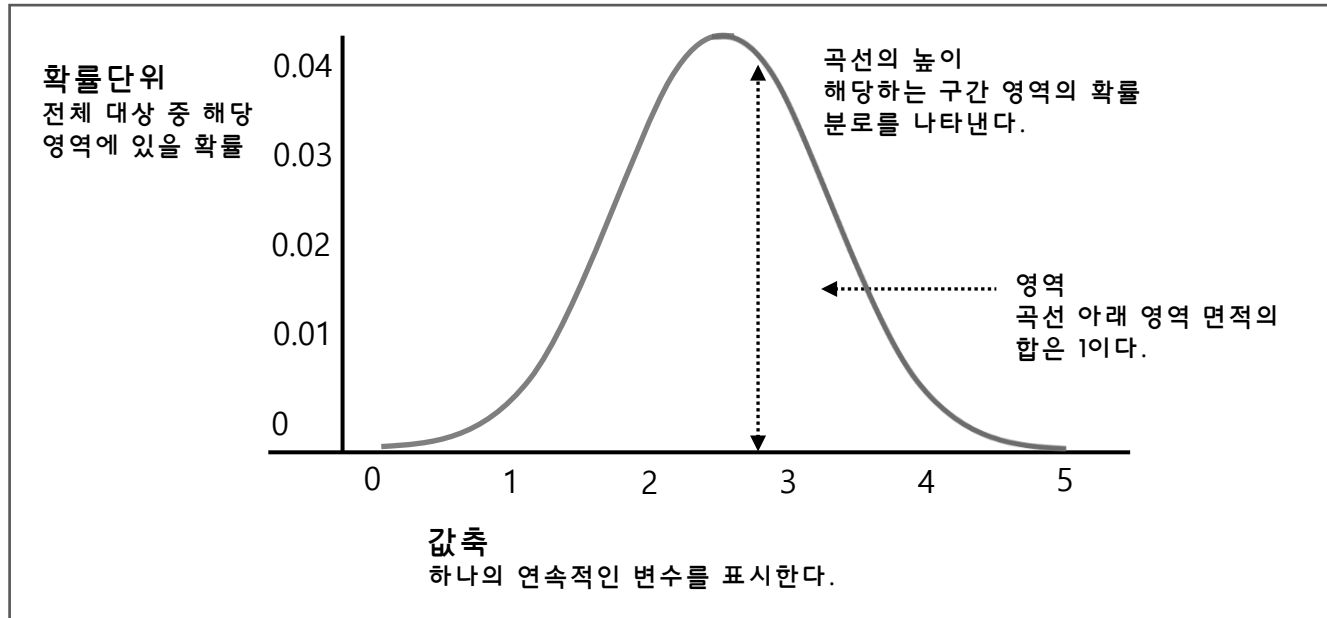
- 막대의 높이는 빈도를 나타냄
- 가로와 세로축은 연속적



## 관계 시각화(7)

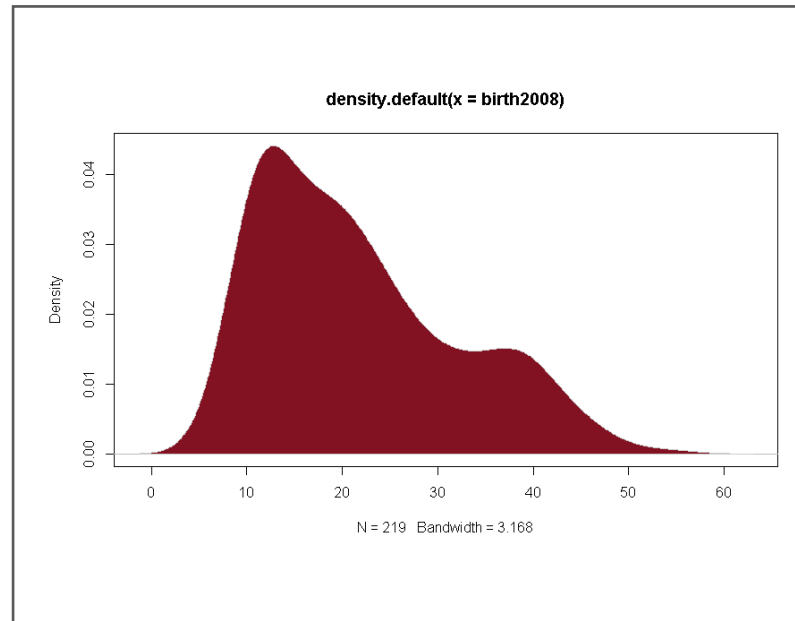
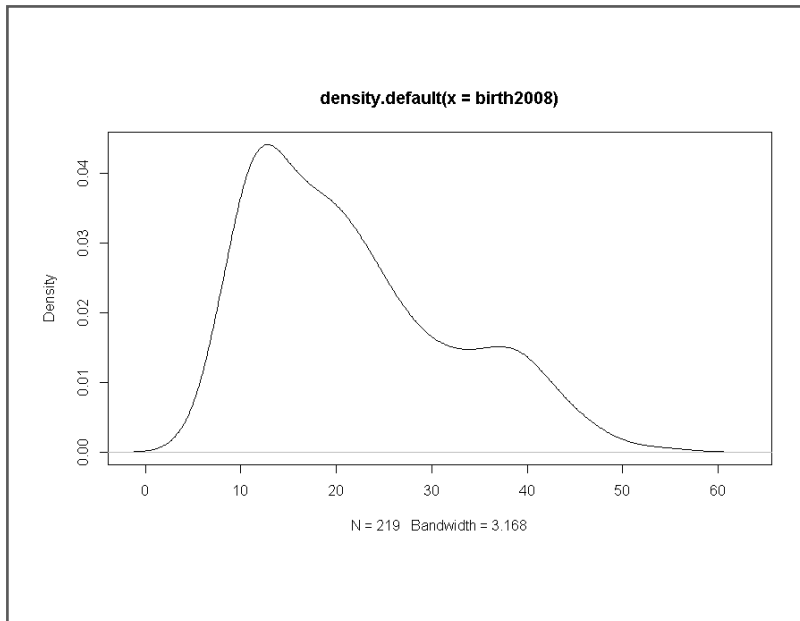
### • 분포-연속 밀도 함수

- 연속적인 확률 분포를 표시
- 면적의 합은 1



## 관계 시각화(7)

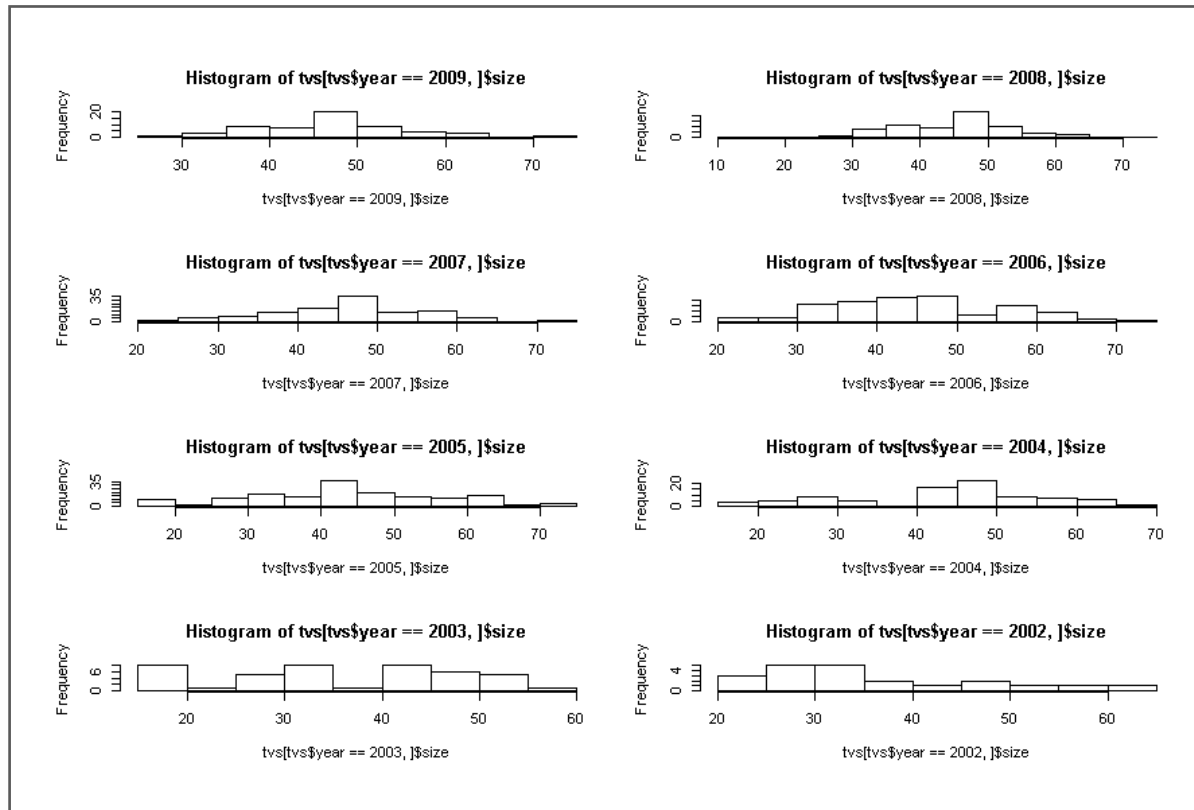
- 분포-연속 밀도 함수



## 관계 시각화(8)

### • 비교-히스토그램 행렬

- 여러 히스토그램을 한 화면에 표시

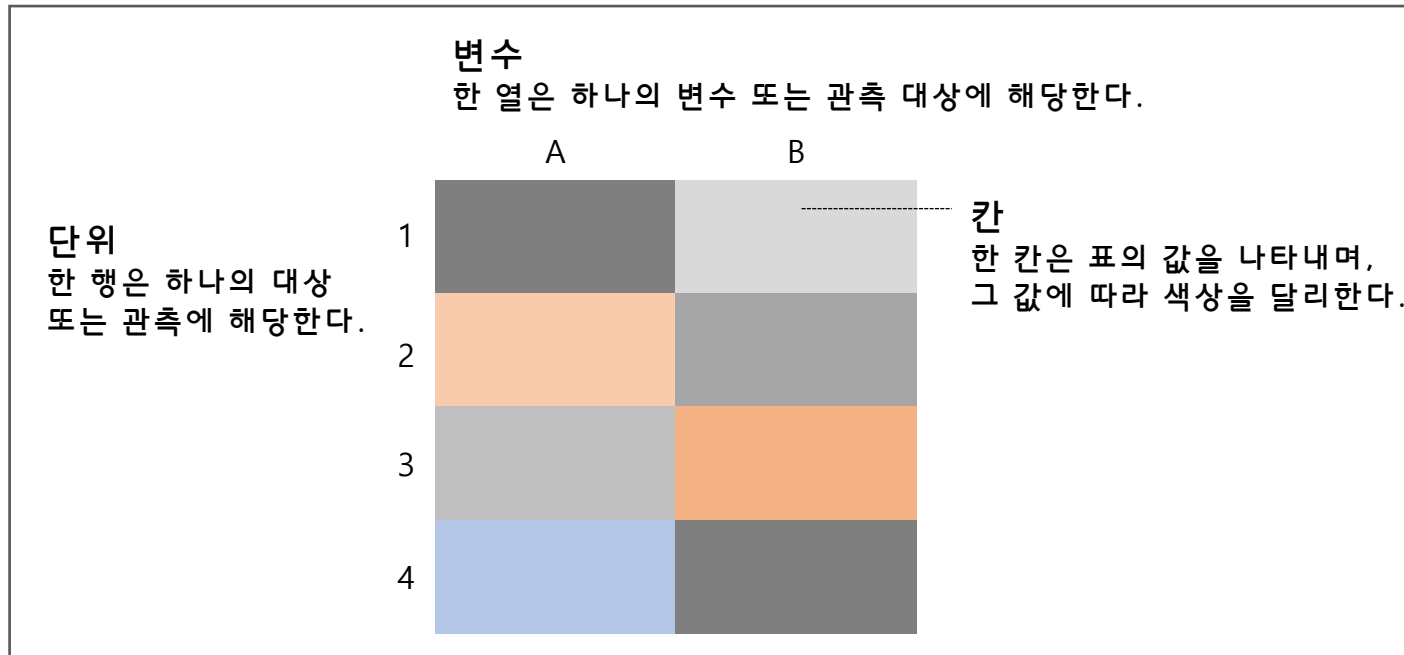


## 비교 시각화(1)

- 비교 시각화란?
  - 다양한 기준으로 집단을 구분해서 찾아내고, 일반 상식에 의거해 아웃라이어를 찾아내는 방법
  - 여러 변수를 비교하고 차원을 축소 한 뒤 아웃 라이어를 찾는 것이 목적
    - 여러 변수의 비교 : 히트 맵, 체르노프페이스, 스타차트, 평행 좌표 그래프
    - 차원 축소: 다차원척도법
    - 아웃라이어 찾기

## 비교 시각화(2)

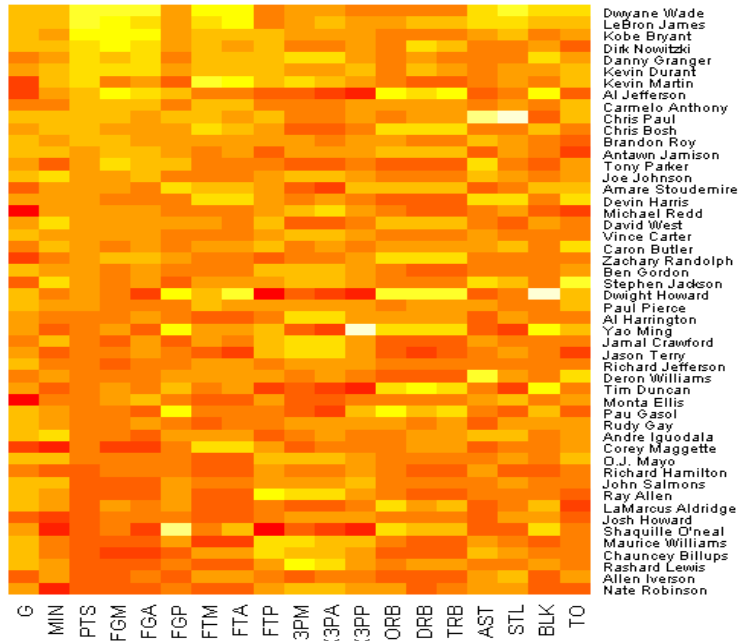
- 여러 변수의 비교-히트 맵
  - 색상으로 값들을 높고 낮은 관계를 한 눈에 파악할 수 있는 표
  - 보통 짙은 색상은 높은 값을, 옅은 색상은 낮은 값을 표현



## 비교 시각화(2)

### • 여러 변수의 비교-히트 맵

	Name	G	MIN	PTS	FGM	FGA	FGP	FTM	FTA	FTP	X3PM	X3PA	X3PP	ORB	DRB	TRB	AST	STL	BLK	TO	PF
1	Dwyane Wade	79	38.6	30.2	10.8	22.0	0.491	7.5	9.8	0.765	1.1	3.5	0.317	1.1	3.9	5.0	7.5	2.2	1.3	3.4	2.3
2	LeBron James	81	37.7	28.4	9.7	19.9	0.489	7.3	9.4	0.780	1.6	4.7	0.344	1.3	6.3	7.6	7.2	1.7	1.1	3.0	1.7
3	Kobe Bryant	82	36.2	26.8	9.8	20.9	0.467	5.9	6.9	0.856	1.4	4.1	0.351	1.1	4.1	5.2	4.9	1.5	0.5	2.6	2.3
4	Dirk Nowitzki	81	37.7	25.9	9.6	20.0	0.479	6.0	6.7	0.890	0.8	2.1	0.359	1.1	7.3	8.4	2.4	0.8	0.8	1.9	2.2
5	Danny Granger	67	36.2	25.8	8.5	19.1	0.447	6.0	6.9	0.878	2.7	6.7	0.404	0.7	4.4	5.1	2.7	1.0	1.4	2.5	3.1





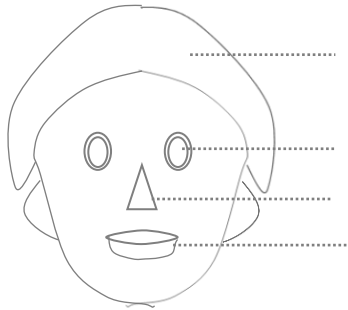
## 비교 시각화(3)

### • 여러 변수의 비교-체르노프페이스

- 다양한 변수를 사람의 얼굴 모양에서 바꿀 수 있는 점으로 달리해서 그리는 방법
- 실생활에서 사람의 얼굴을 쉽게 구분한다는 점에 착안
- 데이터 표현에 따라 달라지는 작은 차이도 쉽게 구분할 수 있다고 가정

#### 얼굴

얼굴 전체로서 하나의 대상,  
또는 데이터의 한 줄을  
나타낸다.

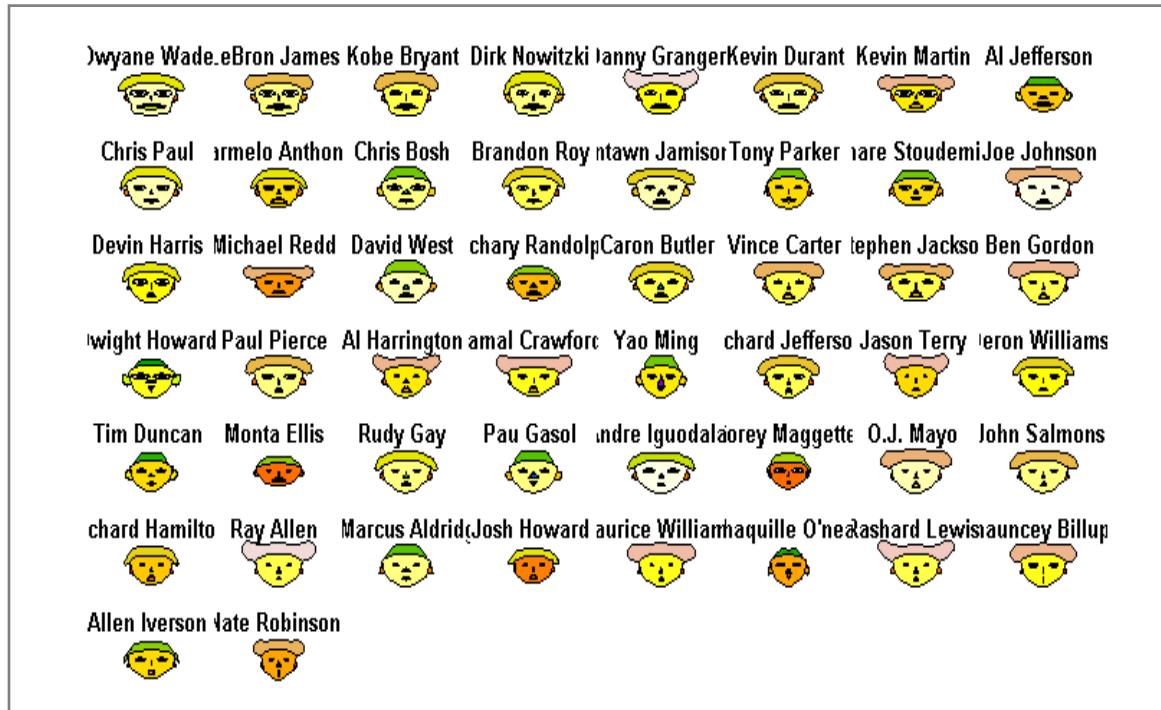


#### 특징

머리 모양의 높이,  
눈과 코의 크기, 입술 모양  
등이 데이터에 따라 다르게  
나타난다.

## 비교 시각화(3)

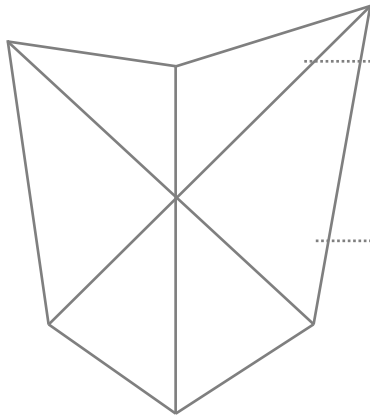
### • 여러 변수의 비교-체르노프페이스



## 비교 시각화(4)

### • 여러 변수의 비교-스타 차트

- 데이터의 수치에 따라 얼굴의 특징을 달리하는 대신, 전체 모양을 바꾸는 방법
- 몇 개의 축을 그리고 전체 공간에서 하나의 변수마다 축위의 중앙으로부터 거리로 수치를 나타냄



#### 돌기

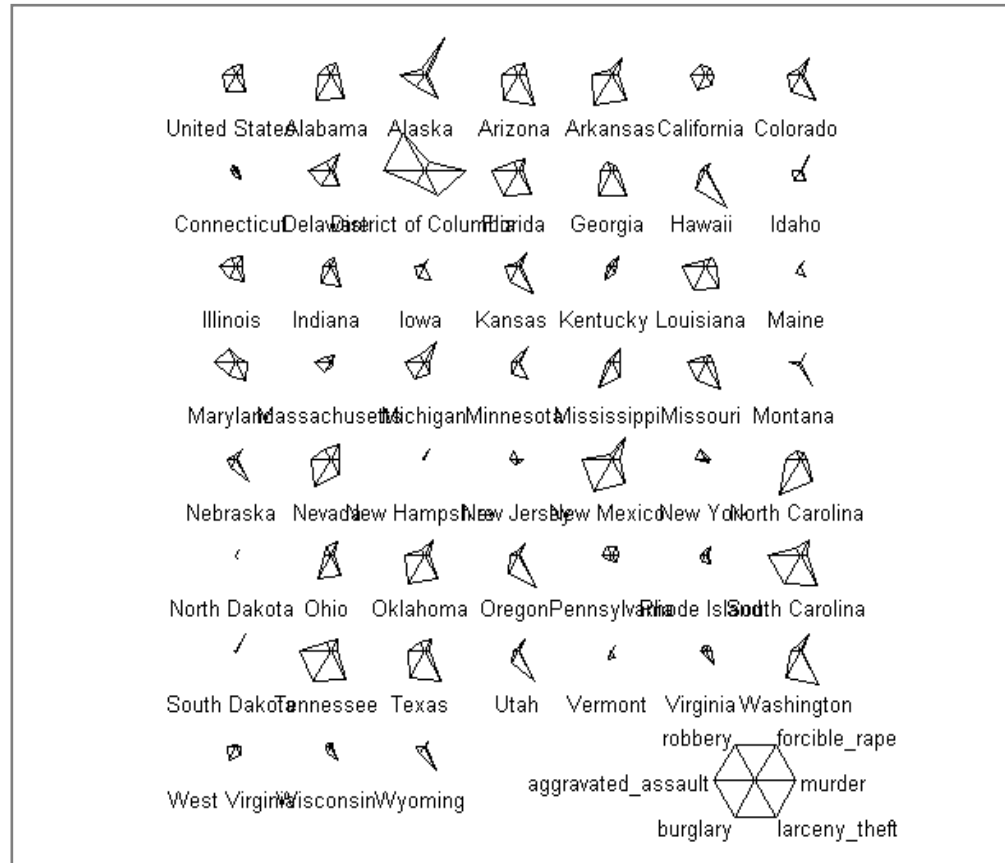
하나의 변수마다 중심으로부터의 거리로 한 변수의 값을 표현한다.

#### 연결선

전체에서 변수들이 서로 어떻게 연관되는지 쉽게 알아보기 위해 그려준다.

## 비교 시각화(4)

### • 여러 변수의 비교-스타 차트



## 공간 시각화(1)

- 공간 시각화란?
  - 지도를 이용한 시각화 기법
  - 지도를 기반으로 하기 때문에 위치영역, 시간과 공간에 따른 차이 및 변화에 대한 것을 다룸
    - 위치 특징 : 위도 및 경도, 지도에 버블로 표현
    - 영역 : 데이터에 따른 색상으로 표현
    - 시간과 공간 : 스몰 멀티플, 애니메이션으로 표현

## 공간 시각화(2)

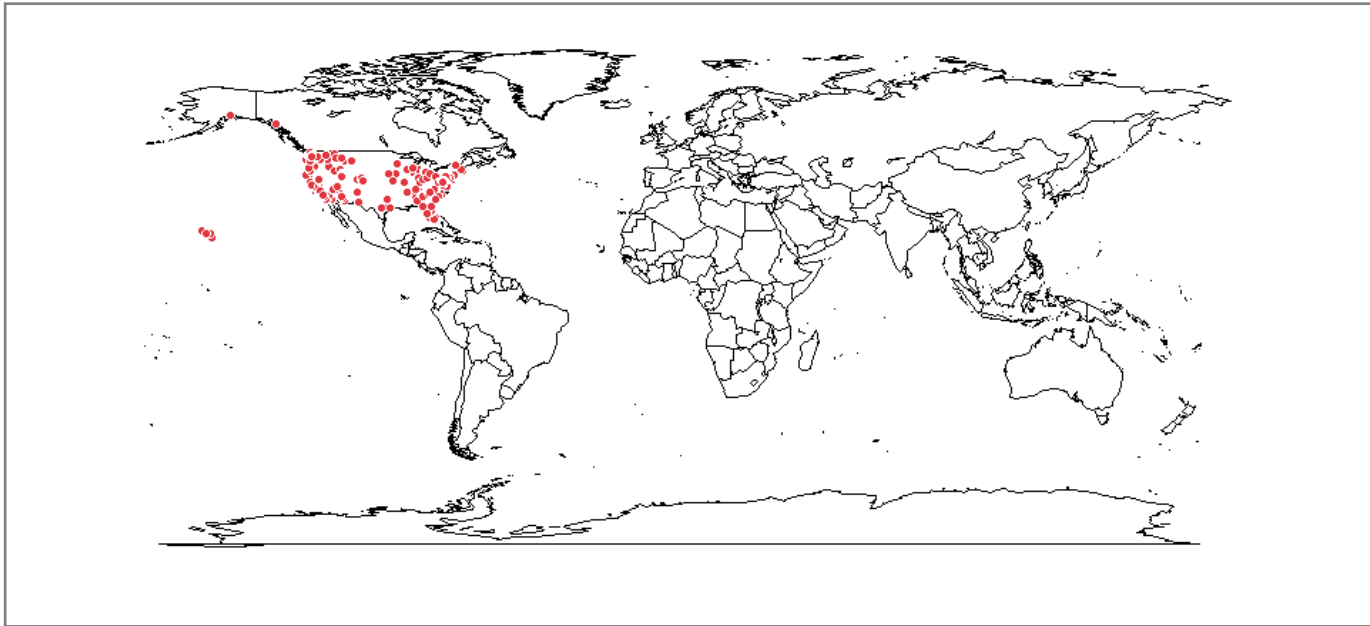
### • 지도 매핑

- 위치의 위도와 경로를 알면 지도에 배치 할 수 있음



## 공간 시각화(2)

- 지도 매핑





## 문제풀이

- 빅데이터 시각화 방법에 대해 설명하고 예시를 들어 서술하시오.
- 파이차트와 누적 막대그래프의 시각화 차이점을 설명하시오.
- 관계 시각화와 비교 시각화에 대한 차이점을 설명하시오.
- 두 개 혹은 다양한 변수에 대해 관계를 비교할 수 있는 시각화를 하기 위해 어떤 그래프를 사용할 수 있는지 작성하시오.

## 요약

- 시간 시각화는 시간에 관련된 시계열 데이터를 시각화 함으로써 트렌드, 경향성을 볼 수 있고, 막대 그래프, 계단 그래프 등의 형태로 시각화 할 수 있다.
- 분포 시각화는 전체의 부분, 시간의 따른 분포에 대해 시각화 함으로써 최대, 최소, 전체 분포에 대해 볼 수 있고, 파이차트, 도넛차트, 평행좌표그래프 등의 형태로 시각화 할 수 있다.
- 관계 시각화는 데이터 간의 상관관계, 분포, 비교에 대해 시각화 함으로써 스캐터 플롯, 스템 플롯, 히스토그램 행렬 구조 등의 형태로 시각화 할 수 있다.
- 비교 시각화는 다양한 기준으로 집단을 구분해서 찾아 차원을 축소한 뒤 아웃라이어를 찾는 것이 목적으로 스타 차트, 히트맵등 다양한 형태로 시각화 할 수 있다.

## 요약

- 새로운 정부 운영의 패러다임인 정부 3.0에 개방된 공공데이터는 정형 또는 쉽게 정형으로 변화 가능한 데이터로 구성되어 있다.
- 표, 시간순 텍스트 등의 다양한 방법으로 시각화 할 수 있다.

## 참고문헌

- [1] Gartner(2011), ‘Hype Cycle for Emerging Technologies, 2011`.
- [2] 김성태, ”새로운 미래는 여는 빅데이터 시대”, 한국정보화진흥원, 2013.
- [3] 허윤, “Revolution R Enterprise 빅데이터 Visualization 분석”, FK BCG Corp. 2013.
- [4] Evelson.B & Yuhanna.N, “The Forrester Wave:Advanced data visualization (ADV) Platforms”, Forrester Research, Q3, 2012
- [5] 네이션 야우, “VISUALIZE THIS”, 에이콘, 2012.
- [6] 이만재, ”공공 데이터와 비주얼라이제이션”, 빅데이터포럼, 2012.
- [7] 이만재, "빅 데이터 접근방식의 공공 데이터 비주얼라이제이션 사례“ 정보와 통신, 2012.9, pp. 36-42.