

---

# 빅데이터의 시각화

## 한국경영정보학회 추계 학술대회

숙명여자대학교 | 시각영상디자인학과  
이지선

# 빅데이터 시각화 I 시각화 이슈

---

- 빅데이터는 정보 시각화에서 중요한 소재중 하나
- 데이터의 풍부함을 포착하기 위한 새로운 방식
  - 대규모 데이터를 탐색하거나 이해할 때 가장 좋은 방법은 시각화
  - “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)
- 시각화에 따라서 빅데이터는 작은 데이터 셋처럼 보일 수도 있음
- 기존의 역할 : 정보전달의 부가적 설명 장치 (기사의 부가적 설명 장치, 단순한 수치 그래프화)
- Big Data 시대 역할
  - 내부 - 방대한 정보를 하나의 Insight로 도출 할 수 있는 시각적 분석 도구(정보전달 및 상황 진단을 위한 주요 프로세스)
  - 외부 - 일반 사용자(소비자)를 위한 새로운 비즈니스 제공

# 빅데이터 시각화 I 시각화 이슈

---

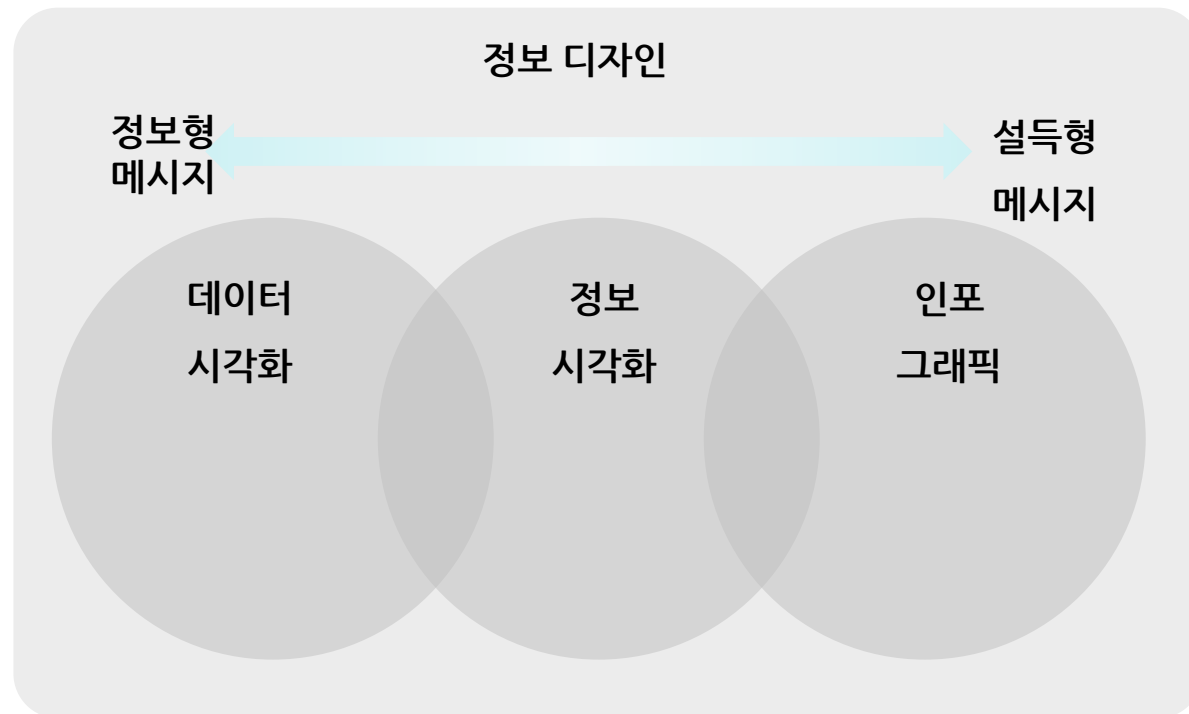
- 빅데이터의 처리기술
  - 데이터마이닝 같은 전통적인 데이터 분석 기술
  - 빅테이블 같은 새롭게 등장하는 빅데이터 처리기법
  - 분석된 데이터를 사용자에게 이해하기 쉽게 시각화하는 기술
- Big Data 분석 도구 상 객관적 정확성이 아직 부족한 상황
- 여전히 Big Data 저장 처리 및 분석 도구 개발 상황이기에 Data Visualization/Inforgraphics에 대한 중요성이 두드러지지 않음

## 국내 데이터 시각화 실태

- 인포그래픽 분야에서 주로 다루어지고 있음
- 데이터 마이닝 작업에 기초한 정보 전달 보다는 메시지 전달을 위한 시각 표현 위주의 작업이 이루어지고 있음

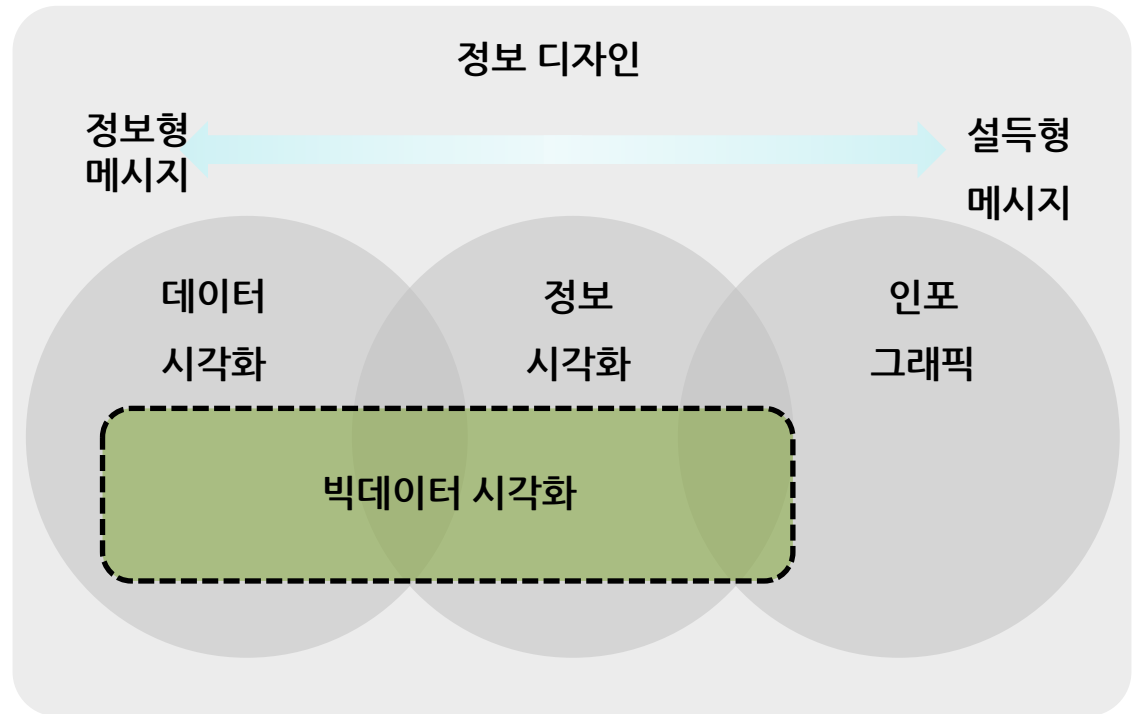
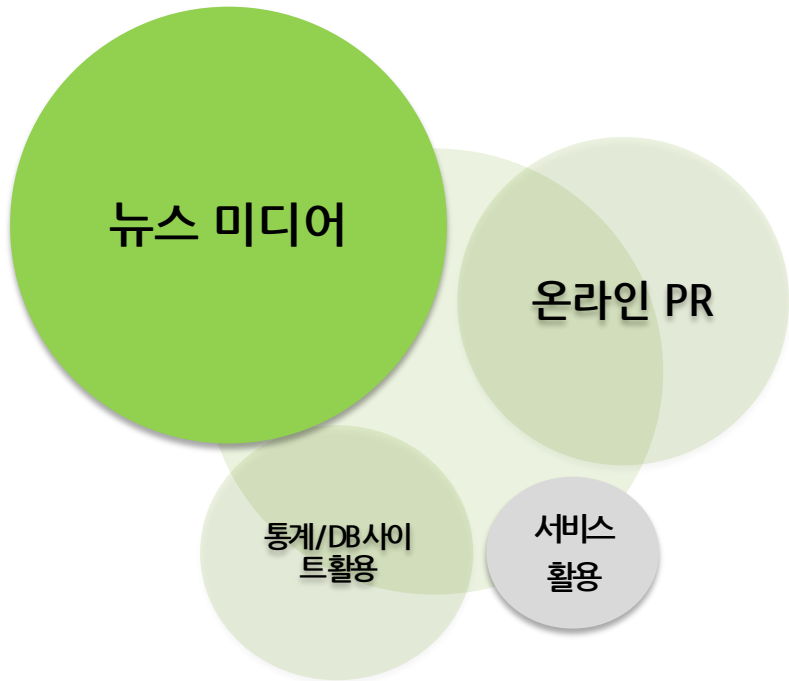
# 시각화 정의 정보 디자인과 시각화

- 데이터 시각화 : 같은 범주 안의 많은 양의 데이터에 의미를 부여하여 효율적으로 전달하기 위한 시각화
- 정보 시각화 : 데이터 시각화 보다 큰 범주에 해당하는 정보들을 시각화
- 정보디자인 & 인포그래픽 : 데이터의 디테일을 나타내기 보다는 보다 그래픽을 적극적으로 이용하여 시각 스토리텔링 형식의 설득형 메시지를 전달



# 빅데이터 시각화 I 시각화 이슈

- Data Visualization/Infographics가 비즈니스에 따라 다양하게 활용되지 못하고 ‘기존의 역할’만 수행하고 있는 상황

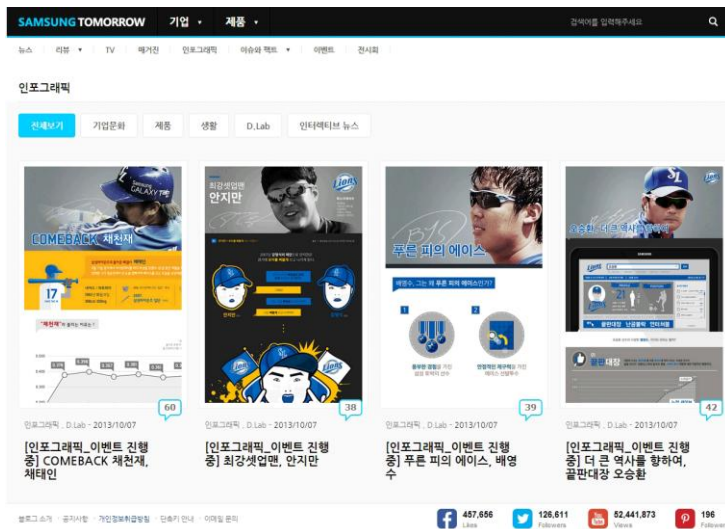


# 빅데이터 시각화 | 역할에 따른 비교

## 디지털 조선



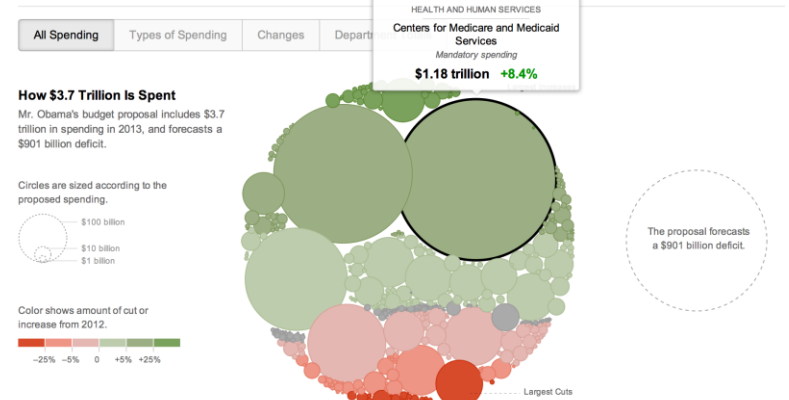
## 삼성 투모로우



## NY Times

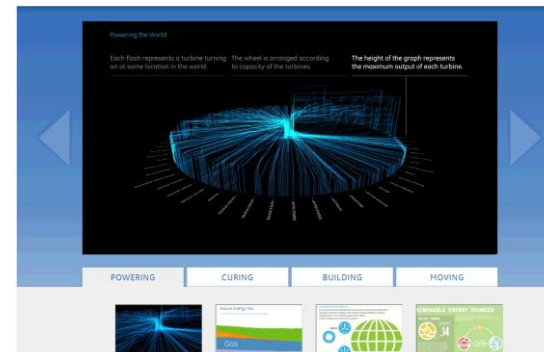
### Four Ways to Slice Obama's 2013 Budget Proposal

Explore every nook and cranny of President Obama's federal budget proposal.



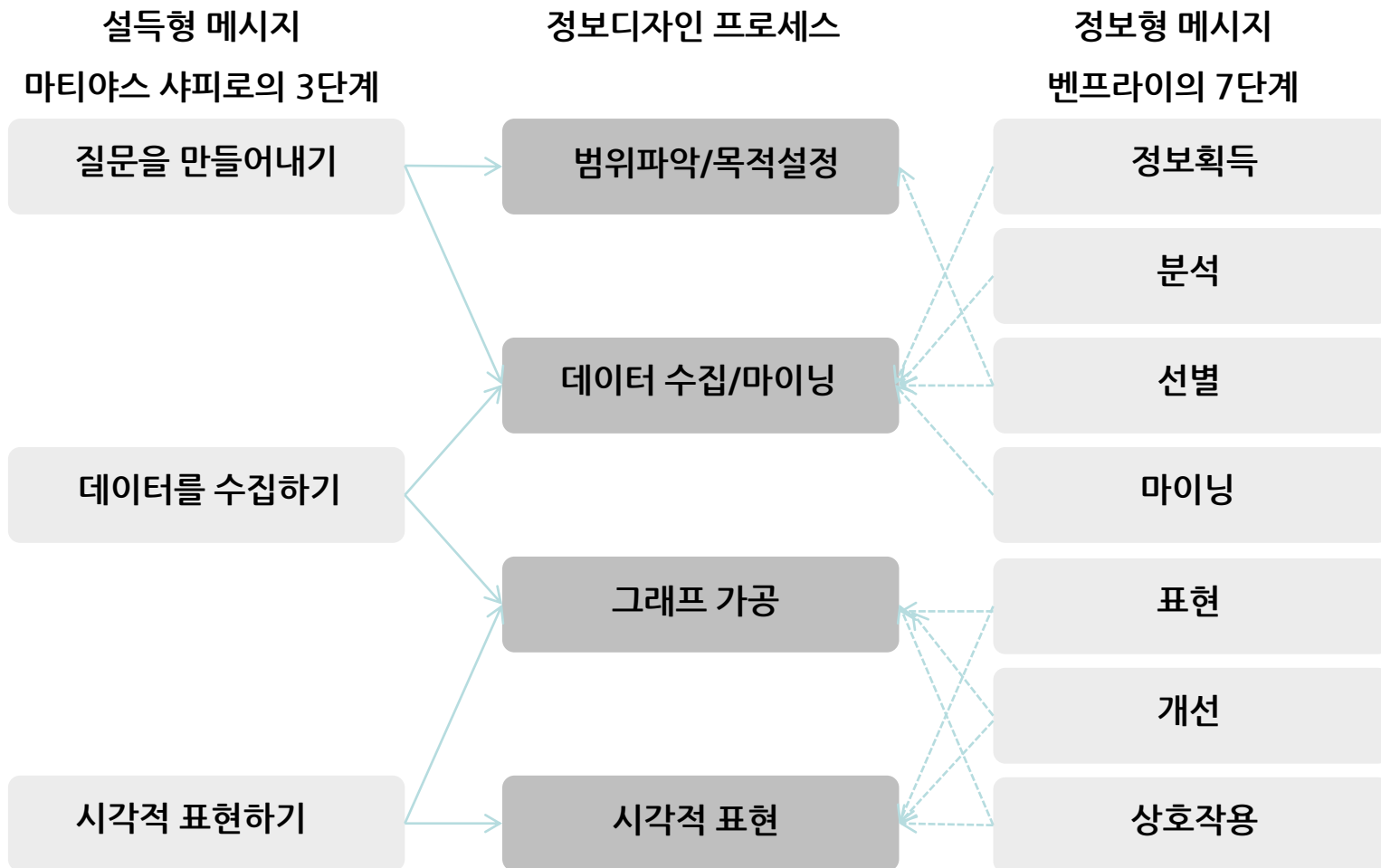
## GE Data Visualization

- GE 의 'How to Show'
- 헬스, 조명, 전기, 소프트웨어 등 GE제품 및 서비스 관련 Big Data Team 구성



# 시각화 방법 시각화 방법 분석

## ● 시각화의 다양한 방법



# 시각화 방법 정보 조직화

- 좋은 정보 디자인
  - 일차적으로 정보를 사용자에게 정확하게 그리고 효율적으로 전달하는 것
  - 정보 디자인은 정보가 많을수록 좋다는 ‘다다익선’의 원칙이 아닌 정보의 필요성과 효율성에 근거해서 선택과 집종의 원칙이 우선
- 정보가 이해하기 쉽도록 조직화되어 배치될 때, 정보의 가치와 유용성은 더욱 증대





# 시각 표현 방법 분석 | 시각 표현 방법

---

- 시각표현 : 데이터에 따라서 변화할 수 있는 일종의 시각적 차원(visual dimension)을 뜻함
- 마티아스 사피로가 분류한 ‘크기’, ‘색상’, ‘위치’, ‘네트워크’, ‘시간’, ‘다중표현기법’을 기준으로 설정
  - 크기: 면적이나 도형의 모양의 확대/축소를 이용하며 사용자가 직관적으로 구별할 수 있기 때문에 가장 많이 쓰이는 표현 중에 하나
  - 색상: 데이터 셋이 많을 때 규칙성과 특이성을 구분해내는데 효과적으로 쓰일 수 있다.
  - 위치: 지도나 가상의 장소와 데이터를 연결하여 나타냄으로써 관찰자가 자신의 정황을 시각화에 투영하여 해석하도록 함
  - 네트워크: 데이터 사이에 관계를 표현하며 각 데이터들을 노드(node)로 연결하는 것과 같으며 사회적 관계를 시각화하거나 방대한 양의 데이터들 사이의 관계를 그룹지어 보여줄 때 유용
  - 시간: 전통적으로 많은 정보디자인에서 쓰였던 방법으로 시간순서에 따라서 데이터가 나열

## 시각화 툴 | 시각화 툴 분류

---

### ● 데이터 마이닝 부터 시각화까지를 제공하는 시각화 툴을 이용

- 기초 수준 : 엑셀, CVS/JSON, 구글 차트 API, Flot, Raphael, D3(Data-Driven Documents), Visual.ly, infogr.am
- 인터랙티브 GUI 컨트롤 : Crossfilter, Tangle,
- 맵핑 : Modest Maps, Leaflet, Polymaps, OpenLayers, Kartograph, CartoDB,
- 전문가수준 : Processing, NodeBox / 전문가 툴 : R, Weka, Gephi  
(출처 : The Top20 data visualization tools)

### ● 각각의 시각화 툴이 시각표현을 구현하는 데 적합한가?

- 빅데이터의 데이터를 마이닝 과정까지 시각표현을 어떻게 구상 할 것인가?
- 마이닝을 통하여 추출된 결과를 원하는 시각표현을 위하여 효율적인 시각화 툴을 이용하였을때 원하는 시각표현을 구현하는것이 가능한가?
- 데이터와 툴을 넘어서선 의도하는 시각표현 메시지를 전달하는 것이 가능한가?

## 결론 빅데이터 시각 표현

- R 등 통계관련 시각화 툴 및 데이터 시각화 프로그래밍 능력과 더불어 시각 표현의 전문성을 갖추는 것이 빅데이터 시각화 질을 향상 시킬 수 있다고 할 수 있음

### 빅데이터 시각화 요건

- 빅데이터에서 다루는 데이터를 어떻게 통합하였는지
- 데이터 고유의 특성이 훼손되지 않으면서 비교되는 통계 품질이 어떠한지
- 의도하는 정보디자인을 위한 통계 데이터의 도출이 적절하게 되었는지 의해 좌우

R 등의  
통계관련  
시각화 툴

데이터 마이닝  
+  
시각화  
프로그래밍 능력

정보디자인의  
시각표현  
전문성

---

이지선

숙명여자대학교 시각영상디자인학과

[jisunlee@sookmyung.ac.kr](mailto:jisunlee@sookmyung.ac.kr)

facebook: jisunlee.net

twitter: jisunlee