
데이터 시각화 - 시각화 디자인

한국 데이터 베이스 진흥원

숙명여자대학교 | 시각영상디자인학과

이지선

데이터 시각화 구현

정보와 시각화

“우리는 정보 과부하나 자료 과다로부터 비롯된 모든 괴로움에 있는 것을 느낀다.

좋은 소식은 그것에 대한 쉬운 해결이 있을지도 모른다는 것이고, 그것은 우리의 눈을 더 사용하는 것이다.”

- David McCandless

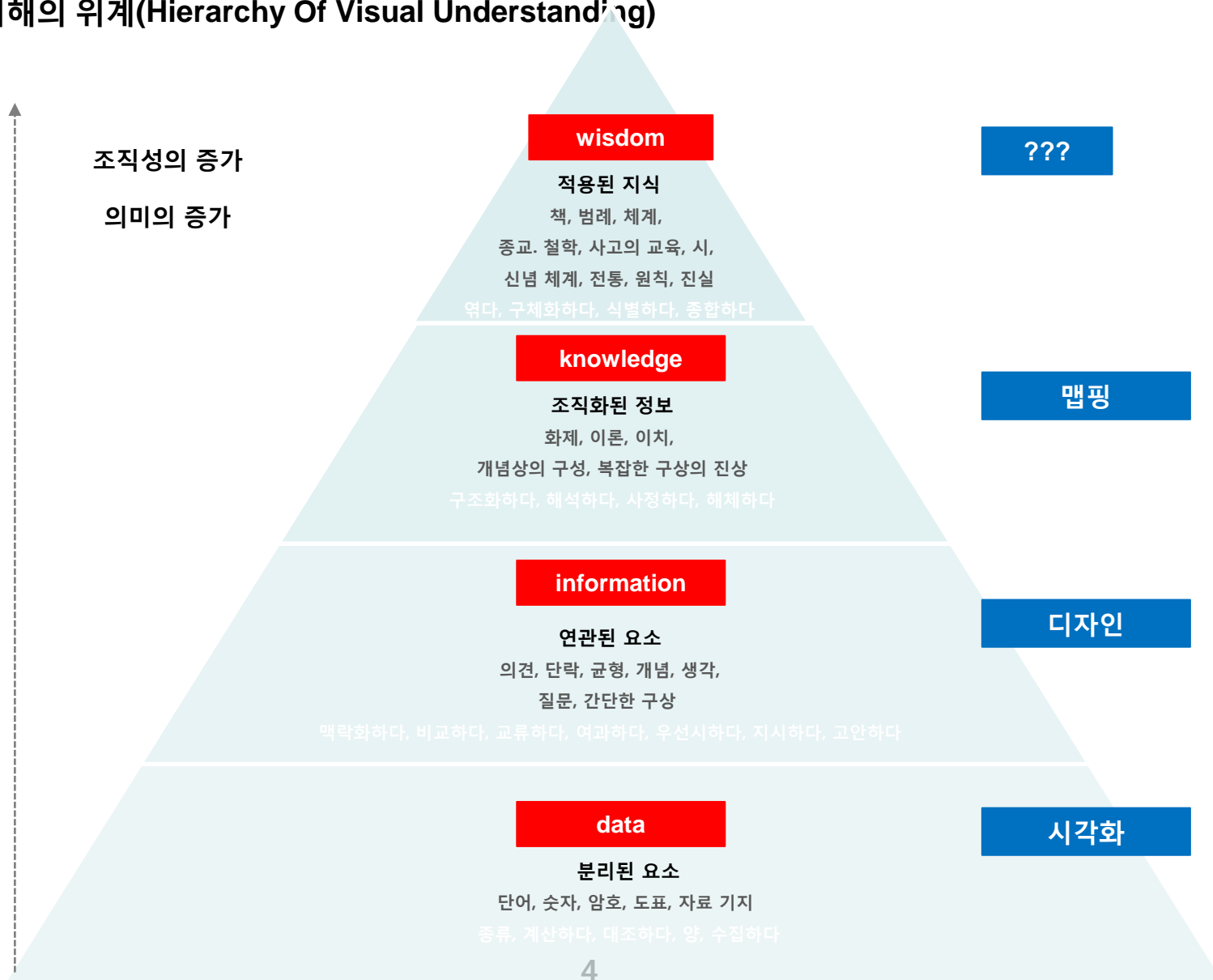
“사람은 보는 것만으로도 데이터에 대해 많은 것을 알 수 있으며 데이터의 의미를 이해하고 그 이야기를 이해하곤 한다.” - 존 터키

“차트와 그래프는 단순히 분석을 위한 도구가 아니라 생각의 소통을 위한 전달체이며, 농담거리를 전달하는 매개이기도 하다.”

- 네이션 야우

시각화 정의 | 시각적 이해의 위계

○ 시각 이해의 위계(Hierarchy Of Visual Understanding)

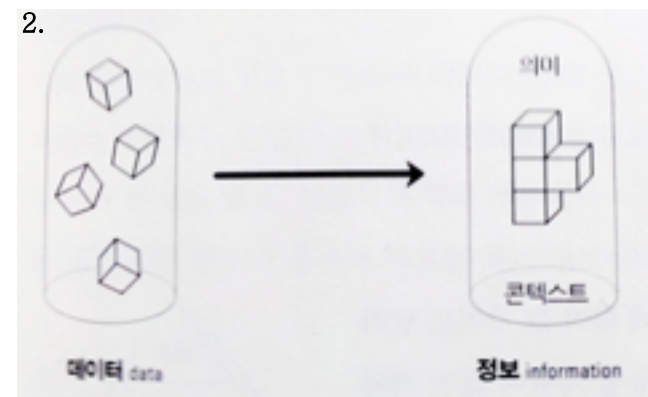
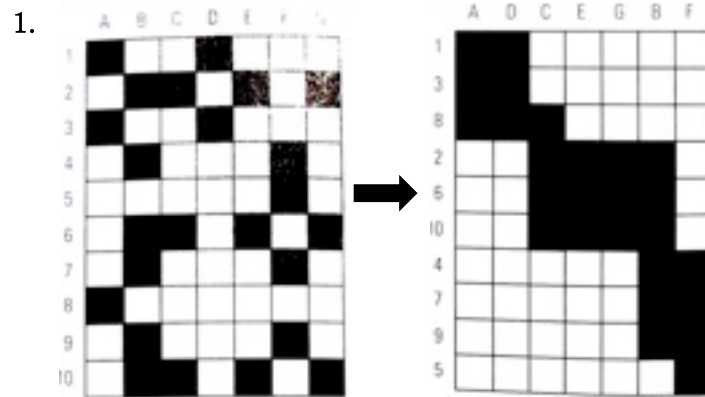


시각화 정의 데이터

- ‘데이터’의 사전적 의미는 재료 또는 자료, 논거를 뜻하는 데이텀의 복수형
- 일반적으로 결론을 내리는 데 근거가 되는 사실이나 참고 자료, 데이터는 정보와 혼용하기도 하고 미디어에 저장된 형태를 일컬음
- 넓은 의미로는 데이터는 연구나 조사, 발견, 수집의 결과인 일종의 기초 자료로서, 정보를 만들기 위한 원자재와 같은 것이라고 정의
- 데이터는 정보 자체가 아니므로 정보로서의 가치가 부족하며, 데이터를 만들어 낸 생산자에게는 유용할 수 있으나 사용자에게 의미를 전달하기에는 적절치 않음
- 가공되지 않고 의미를 갖지 않은 상태의 개체
 - 데이터 디자인이라는 표현을 쓰지 않는 것에서 알 수 있듯이, 데이터가 분석의 대상이 될 수도 있어도 그 자체가 우리가 디자인하려는 대상은 아님
- 효율적인 정보 전달에서는 데이터를 직접 보여 주지 않음 : 그 이유는 불완전하고 비연속적이며 완전한 메시지가 아니므로 정보 전달 측면에서 가치가 없기 때문

시각화 정의 정보

- 정보는 데이터와 달리 그 자체만으로 의미가 있으며, 생산자와 사용자의 관점에 따라 다르게 전달될 수 있으며, 형태와 형식을 가짐
- 데이터와 정보를 더 명확하게 구분하려면 정보 생산과 활용 과정에서 컨텍스트를 고려해야 함, 컨텍스트가 없으면 정보가 존재할 수 없음
 - 컨텍스트는 데이터의 환경과 관계될 뿐 아니라 그것이 어디에서 왔는지, 왜 소통되어야 하는지, 어떻게 배열되는지, 이해하는 사람의 태도와 환경이 무엇인지를 의미
 - 정보는 사용의 주체나 상황에 따라 의미와 가치가 다르기 때문에 정보가 생산되고 사용되는 컨텍스트가 중요



1. Richard Saul Wurman, [Information Anxiety 2], Que, 2000, P.28

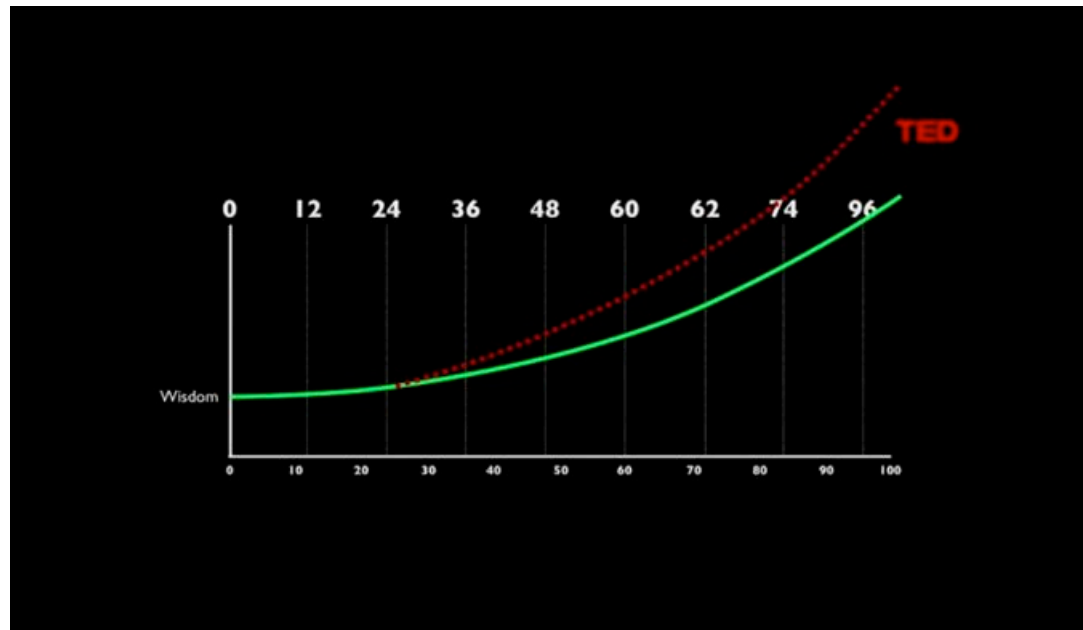
2. 데이터의 재배열 데이터를 의미 있는 패턴으로 조직화 하여 정보로서 가치가 생성되도록 하는 것

시각화 정의 지식

- 지식은 인간이 생활을 영위하면서 인위적으로 습득하는 고도의 논리적 상식이자, 정보의 상위 개념이며 모든 경험의 산물
- 경험을 통해 다른 관점과 방법으로 해석 가능
 - 경험을 통해 형성된 지식은 특정한 세부 사항만을 설명하는 것이 아니라 다양한 상황에서 적용할 수 있게 일반화된 것을 의미
- **스토리 : 지식을 전달하는 데 가장 효과적인 방법**따라서 정보의 조직화에서도 스토리텔링 개념이 중요하게 적용된다.
 - 좋은 스토리는 섬세하게 묘사되는 풍부한 세부성과 경험을 제공하는 서사성이 있고 다양한 해석을 허락
 - 상대를 보면서 이야기하는 스토리는 듣는 사람의 흥미와 경험에 따라 변화하고 정보를 더욱 개인적 수준으로 만듦

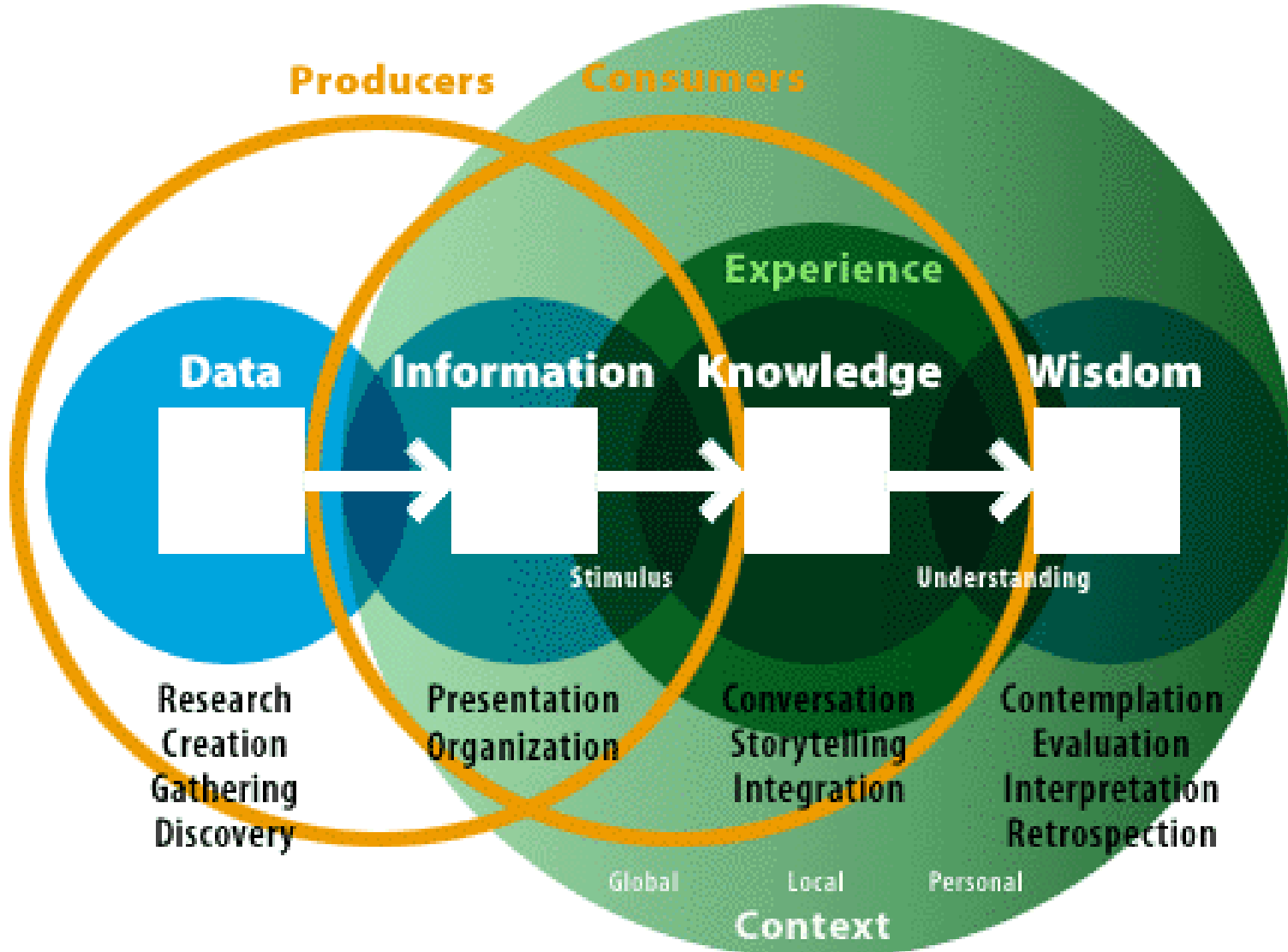
시각화 정의 지혜

- 지혜는 고차원 방법으로 우리가 사용하기에 충분하고도 이상적인 패턴을 이해하는 정보의 최종 단계
- 지혜는 개인적 이해의 수준에 따라 결정되는 것으로 도달하기 어려운 단계
- 지혜는 정보와 지식의 개인화에 의해 생성되며, 인위적으로 전달하거나 공유할 수 있는 것이 아님
- 지혜는 다른 단계들보다 훨씬 추상적이고 철학적인 단계



시각화 정의 | 종합

- 정보 인터랙션 디자인(Information Interaction Design: A Unified Field Theory of Design)



시각화 정의 데이터 시각화

○ 데이터 시각화(Data Visualization)

- 데이터의 시각적 표현의 연구영역을 말하며, 이는 도식적 형태안에 추상적으로 표현된, 속성(attributes)이나 변수(variables)를 가진 단위를 포함한 정보를 이야기함
- 관련 분야 : 인포그래픽(info graphics), 정보시각화(information visualization), Scientific visualization, statistical graphics
- 데이터 시각화 범위: 마인드맵, 뉴스 표현, 데이터 표현, 관계들의 표현, 웹사이트들의 표현, 기사와 리소스들, 툴과 서비스
- 데이터 시각화 주요 분야: 통계적 그래픽(statistical graphics), 주제 지도학(thematic cartography) - 프랜들리(2008)

시각화 정의 정보 시각화

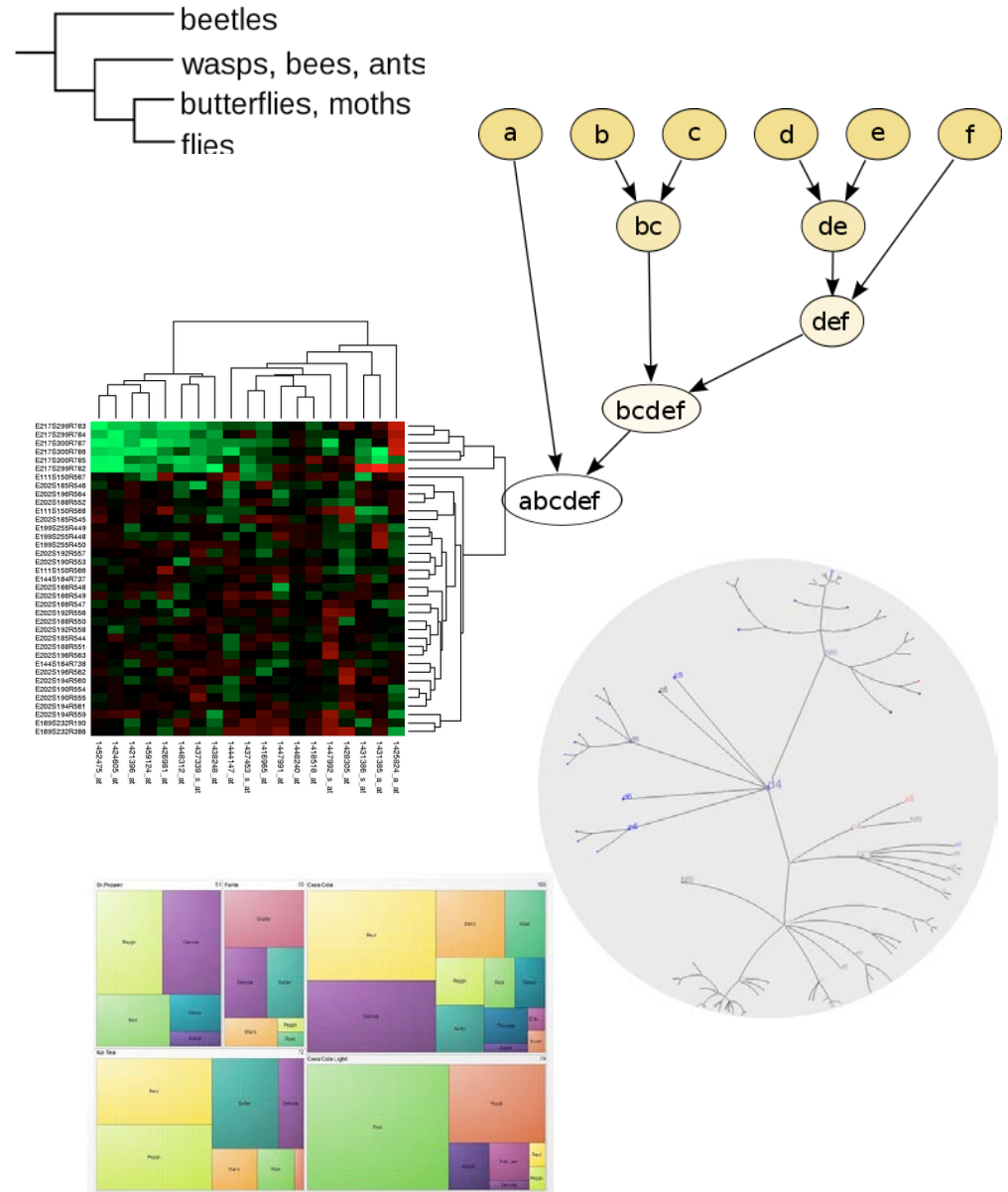
○ 정보 시각화(information visualization)

- 소프트웨어 시스템, 라이브러리, 서지 데이터베이스에서의 코드의 라인들과 화일들 같은 비수치 정보, 인터넷의 네트워크관계 등 큰 범위의 집합들에 대한 시각적 표현방법의 간학문 연구영역
- 방대한 양의 정보를 한번에 사용자가 보고, 탐험하고 이해하도록 시각적 표현방법과 인터렉션 기술들을 이용하여 추상적 정보를 직관적 방법으로 전달하기 위한 접근방법을 창조하는 것에 초점을 맞춤
- HCI, 컴퓨터공학, 그래픽, 시각디자인, 심리학, 비지니스 방법론등의 연구영역으로 부터 발전했으며 과학적 연구, 디지털 라이브러리, 데이터 마이닝, 금융데이터 분석, 시장조사, 제품생산 컨트롤, 약품개발 등에서 중요한 부분에 적용이 증가

시각화 정의 정보 시각화

정보 시각화 방법 및 테크닉

- 분기도(Cladogram)
- 수지도(Dendrogram)
- 정보시각화 레퍼런스 모델
- 그래프 드로잉
- 히트맵(heat map)
- 쌍곡선트리(Hyperbolic tree)
- Multidimensional scaling
- Problem Solving Environment
- 트리매핑(Treemapping)

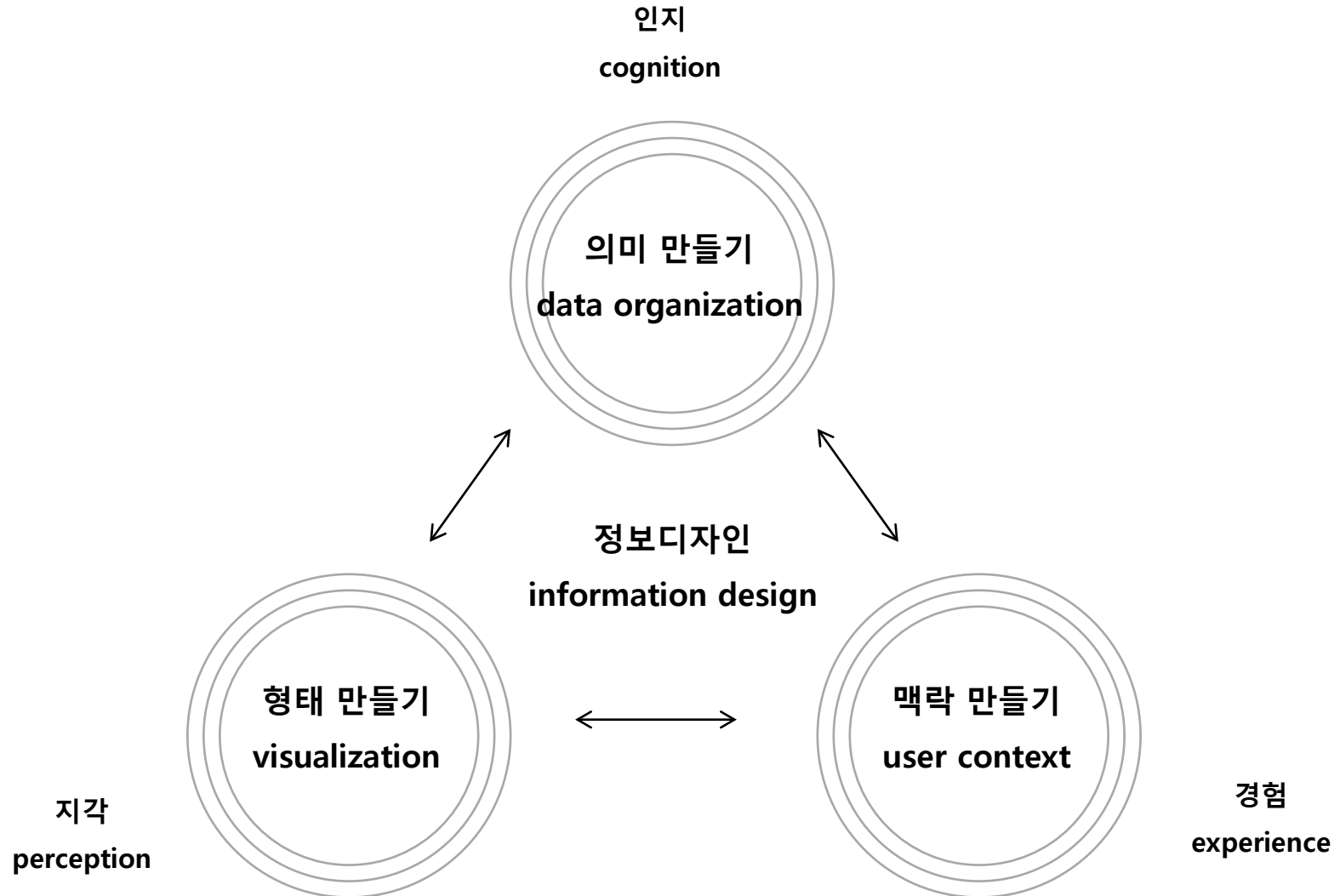


시각화 정의 정보디자인

- 정보 디자인(information Design)
 - 사람들이 사용할 수 있는 효과적인 효율적인 정보 준비의 연습과 기술 데이터가 복잡하고 구조적이지 않은 것에 대하여, 시각적 표현 방법은 좀더 보는 사람들이 좀더 명확하게 의미할 수 있도록 표현하는데 도움
 - 사람들이 정보에 쉽게 접근하고 효율적으로 사용할 수 있도록 하는 것
 - 수집한 데이터를 조직화하고 시각화하며, 사용자의 경험과 전달 수단에 관한 사용 콘텍스트의 과정을 창조
- 정보 디자인 의미 만들기 : 정보의 조직화에 의한 사용자의 정보 해석 과정으로 인지와 관련
- 정보 디자인 형태 만들기 : 정보 시각화에 의해 사용자의 감각을 통한 외부 자극을 수용하는 과정으로 지각과 관련
- 정보 디자인 콘텍스트 만들기 : 사용자의 정보 사용 정황에 대해 정보 디자인의 모든 과정에서 고려하는 것으로 사용자의 경험과 관련

시각화 정의 정보디자인

- 정보디자인 개념도



시각화 정의 인포그래픽

- 인포 그래픽(information graphics, infographic)
 - 정보, 데이터 또는 지식등의 그래픽 정보 표현방법
 - 패턴이나 트렌드를 보기 위하여 인간의 시각능력을 강화하는 그래픽을 이용하여 인지를 강화시킴
 - 사인, 맵, 저널리즘, 테크니컬 라이팅, 교육 등 복잡한 정보를 빨리 명확하게 표현
 - 소셜 큐레이션의 수단으로 이용



시각화 정의 메시지로서의 정보

○ 정보형 메시지

- 정보 자체만을 전달하여 사용자의 이해와 활용에 초점을 맞춤
- 정확한 사실을 전달해야 하므로 장식적 요소와 같은 부가적 표현 없이 **정보 전달의 기능성을 강조**

○ 설득형 메시지

- 정보 전달과 사용자의 이해를 넘어 태도나 행동의 변화를 유도
- 정보 사용자 관심을 끌기 위한 설득적 표현 부가 요소 가미, 과장과 의외성, 이야기와 오락 등 **재미적 요소**로 구성

○ 정보형과 설득형의 통합 메시지

- 전달 과정에서 사용자의 흥미를 유발하여 정보의 효율성을 높이는 **인포테인먼트** 강조

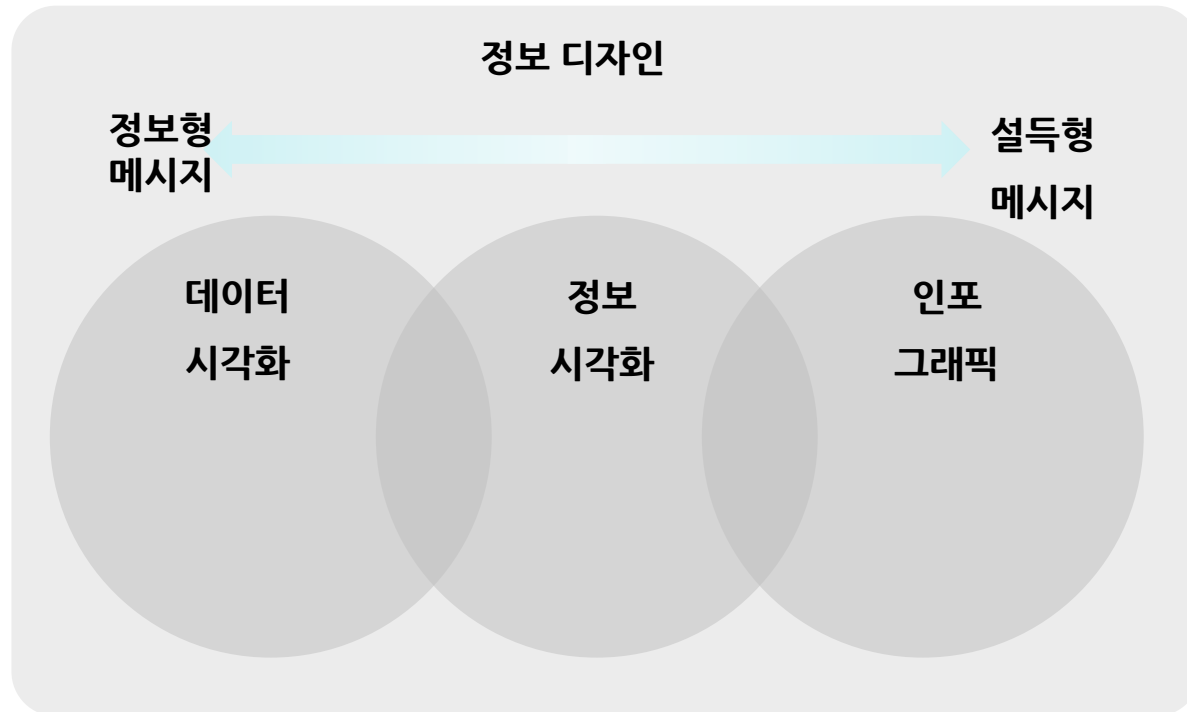


1. 정보형메시지 서울지하철노선도

2. 미국 정부의 자산, 부채 수치 정보와 시각,청각 정보의 기억 정도를 모션 그래픽으로 표현

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

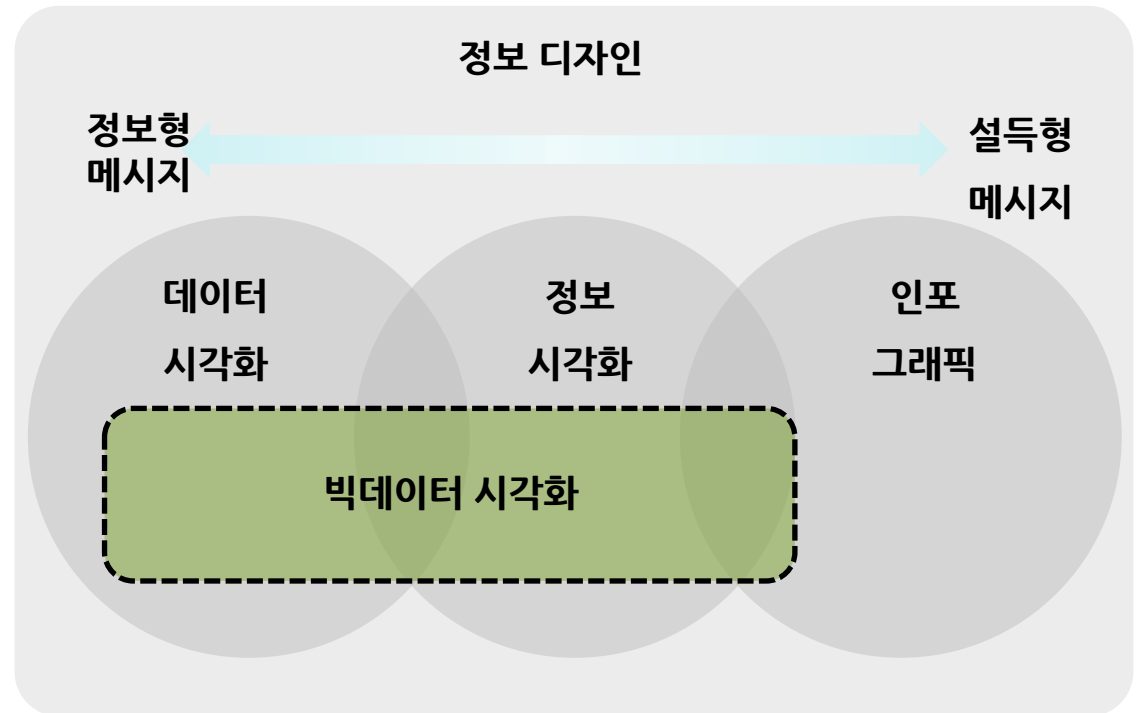
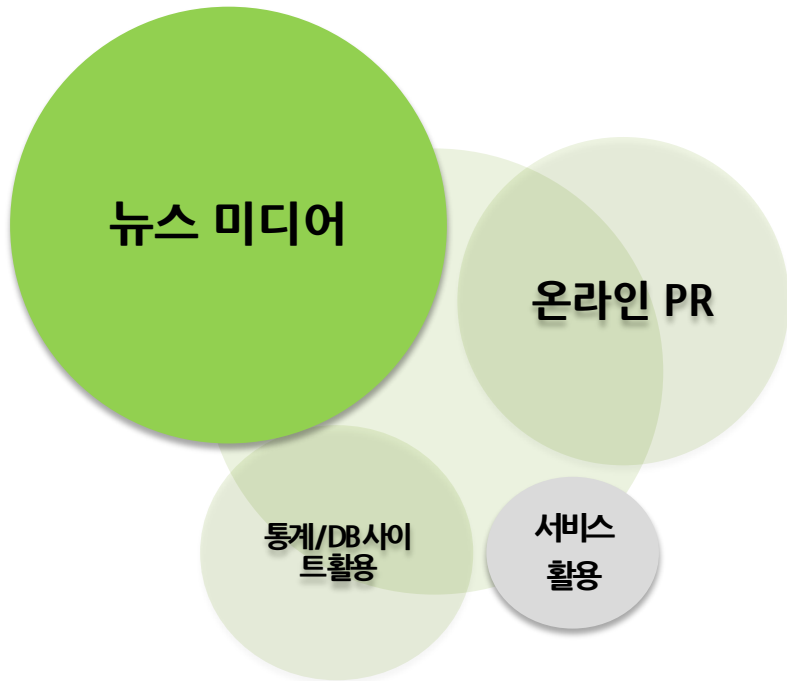
- 빅데이터의 처리기술
 - 데이터마이닝 같은 전통적인 데이터 분석 기술
 - 빅테이블 같은 새롭게 등장하는 빅데이터 처리기법
 - 분석된 데이터를 사용자에게 이해하기 쉽게 시각화하는 기술
- Big Data 분석 도구 상 객관적 정확성이 아직 부족한 상황
- 여전히 Big Data 저장 처리 및 분석 도구 개발 상황이기에 Data Visualization/Infographics에 대한 중요성이 두드러지지 않음

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

데이터 시각화 I 시각화 이슈

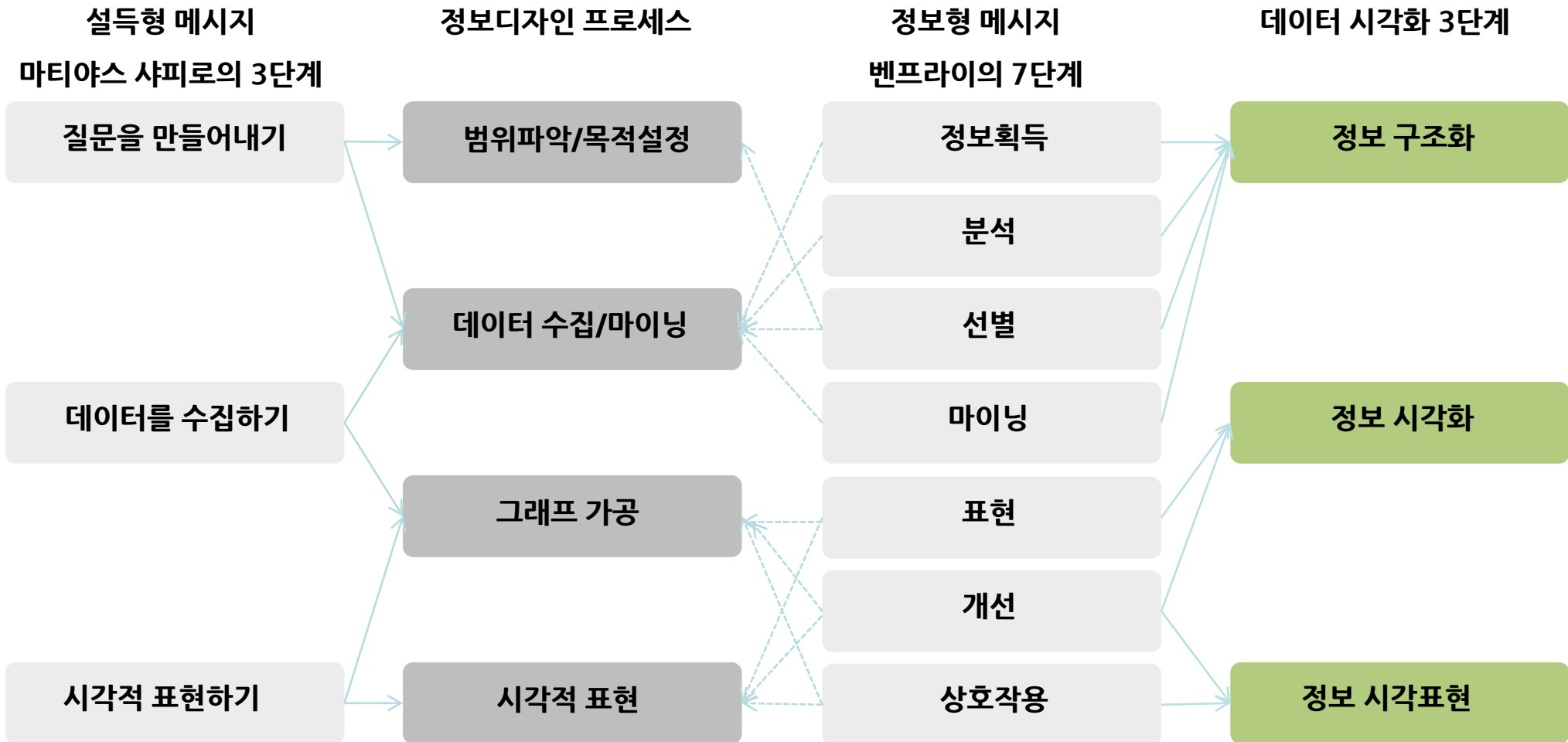
- Data Visualization/Infographics가 비즈니스에 따라 다양하게 활용되지 못하고 '기존의 역할'만 수행하고 있는 상황



관련 주요 사례

시각화 프로세스 프로세스 분석

시각화의 다양한 방법



시각화 방법 | 시각화 3단계 방법

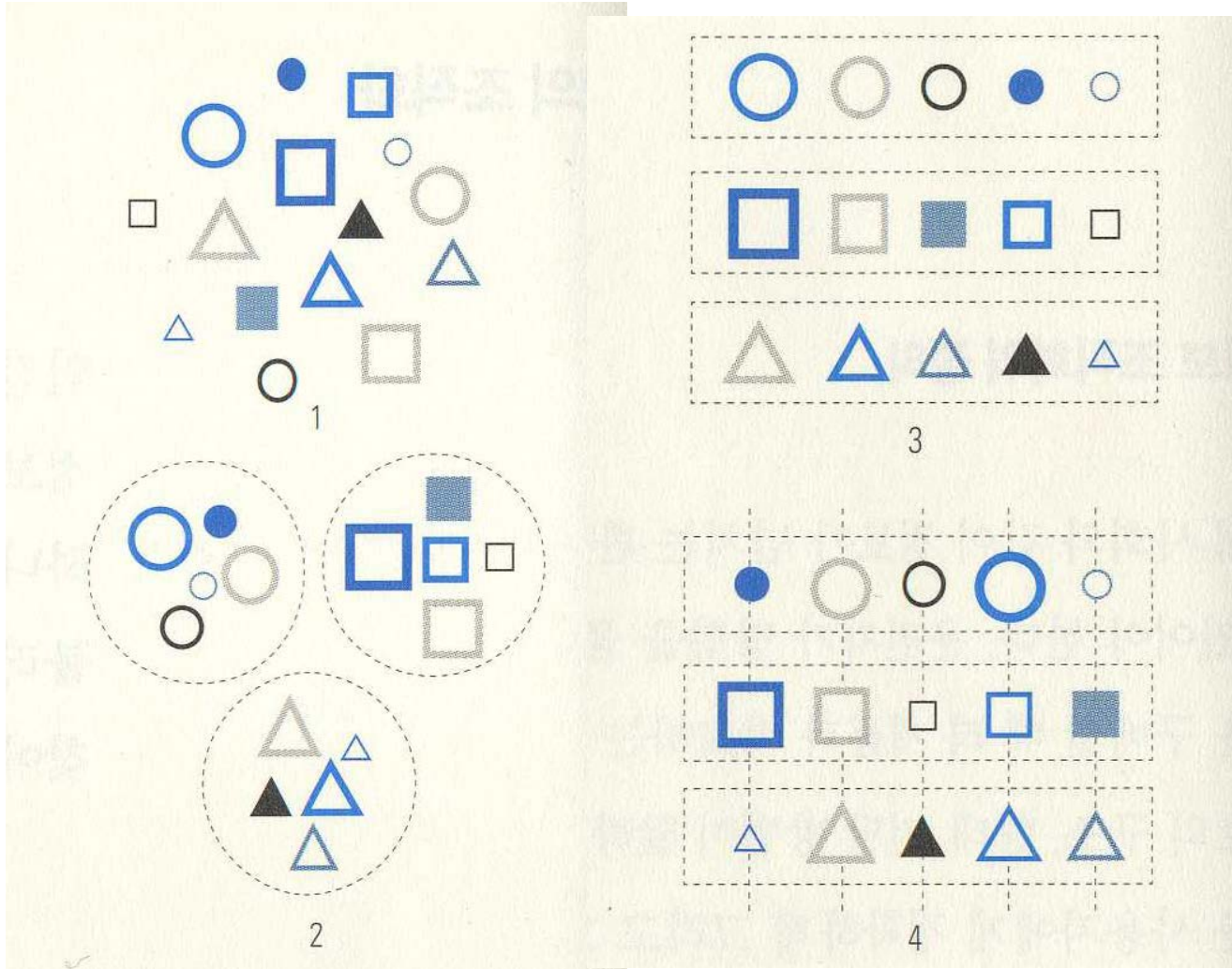
- 정보구조화, 정보시각화, 정보 시각표현의 3단계 시각화 방법

시각화 방법



시각화 방법 | 1단계 정보 구조화

● 정보 구조화



1. 데이터 data : 자료 또는 논거
2. 분류하기 classifying : 분류하여 유사한 것끼리 묶는다.
3. 배열하기 arranging : 값의 의미에 따라 배치한다. 리처드 솔 워먼(Richard Saul Wurman)의 LATCH 방법
4. 관계맺기 organizing : 데이터 사이의 관계를 만든다.

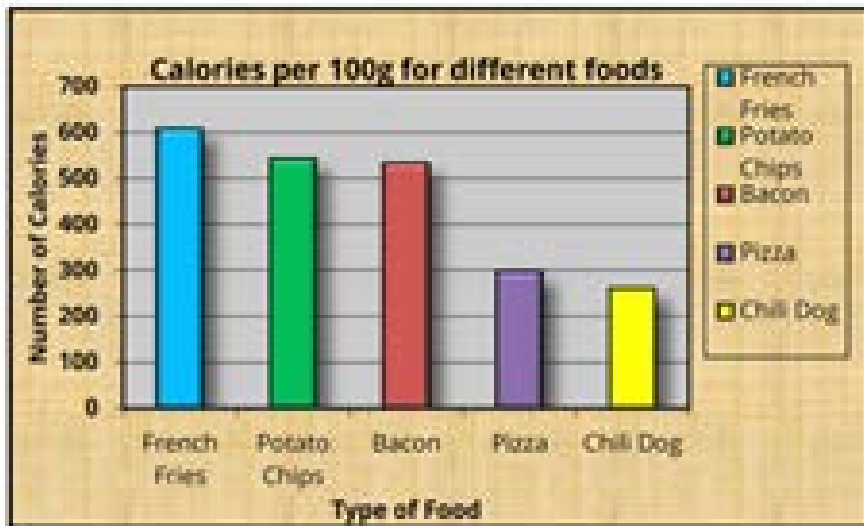
시각화 방법 | 2단계 정보 시각화

- 네이션 야우(Nathan Yau)의 『비주얼라이즈 디스(Visualize This)』에서 제시된 차트와 그래프의 쓰임새를 바탕으로 정리한 정보 시각화 방법
- 대체로 시각화 틀에서 제공하는 그래프의 종류에 따라 시각화의 한계가 정해짐

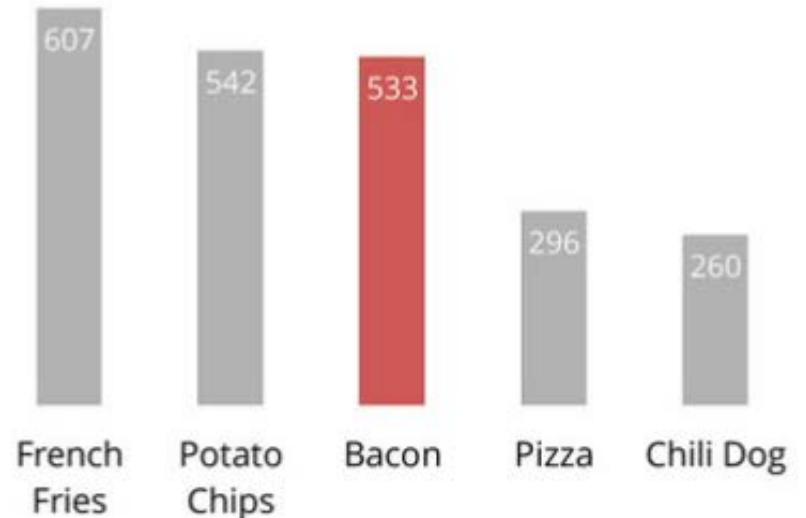
정보 시각화 방법				
시간 시각화	분포 시각화	관계 시각화	비교 시각화	공간 시각화
막대그래프 누적 막대그래프 점그래프	파이차트 도우넛차트 트리맵 누적연속그래프	스캐터 플롯 버블차트 히스토그램	히트맵 체르노프 페이스 스타차트 평행 좌표계 다차원 척도법	지도 매핑

시각화 방법 | 3단계 정보 시각 표현

- 에드워드 터프티(Edward Tufte)가 저술한 『The Visual Display of Quantitative Information』에서는 데이터 링크를 더 이상 지울 수 없는(non-erasable) 그래픽의 핵심이며, 숫자의 변화를 비중복(non-redundant)적으로 표현하는 잉크라고 소개
- 데이터링크가 아닌 것과 중복되는 데이터링크를 제거해 데이터링크 비율을 올리는 것이 데이터를 그래픽 디자인으로 올바르게 표현하는 방법
- 이를 위해서는 정보 디자인 요소를 강조하기 위한 그래픽의 7요소 및 타이포그래피, 그리드, 색상, 아이소타이프 등의 그래픽 기본원리를 이해하는 것이 중요



Calories per 100g



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?

- 데이터를 마이닝 과정까지 시각표현을 어떻게 구상 할 것인가?
- 마이닝을 통하여 추출된 결과를 원하는 시각표현을 위하여 효율적인 시각화 툴을 이용하였을 때 원하는 시각표현을 구현하는 것이 가능한가?
- 데이터와 툴을 넘어선 의도하는 시각표현 메시지를 전달하는 것이 가능한가?

데이터 시각화 사례 | 결과물 형태에 따른 분류

- 데이터 시각화를 결과물 형태에 따라서 2D, 모션 영상, 인터랙션의 출력 결과물이라는 세 가지 분류로 나누고 결과물을 분석

2D 이미지	인쇄물, 온라인 이미지
모션 영상	모션 인포그래픽, 데이터 시각화 동영상
인터랙티브	인터랙티브 웹, 인터랙티브 앱

- 빅데이터가 갖고 있는 방대한 데이터 양을 보여주기 위해 대부분 인터랙티브 웹을 사용했음
- 정보의 시각적 분석을 위한 툴(Visual Analytics Tool) : IBM 매니 아이즈(Many Eyes), 타블로 퍼블릭(TableauPublic), 임퓨어 R(Impure R), 데이터 히어로(Data Hero), 쿼드리그램(Quadrigram) 등
- 데이터를 웹상에서 처리하고 분석해 결과물을 공유 할 수 있는 형태로 템플릿을 제공함으로써 데이터 시각화에 사전지식이 없더라도, 정보의 형식과 종류만 잘 알고 있다면 원하는 결과를 만들어 낼 수 있음

사례적용

○ 한국도로공사 OpenOASIS의 빅데이터 시각화 개발

- 고속도로 교통자료를 개방하여, 창의적이고 실현가능한 아이디어를 발굴하고 한국도로공사가 보유하고 있는 공공 DB의 활용도를 극대화
- OASIS(Operations Analysis and Supportive Information System)는 고속도로 모든 교통이력자료(실시간 자료가 아닌 Historical Data 즉, 하루 지난 어제까지의 데이터)를 저장하는 고속도로 DB시스템의 고유명칭
- 제 2회 **OpenOASIS 교통자료 활용 공모전**
https://docs.google.com/forms/d/1Rc9X0XSjZEF6yYQrUUhiY_MXQLSSfs3FKPdaq6NBX4A/viewform

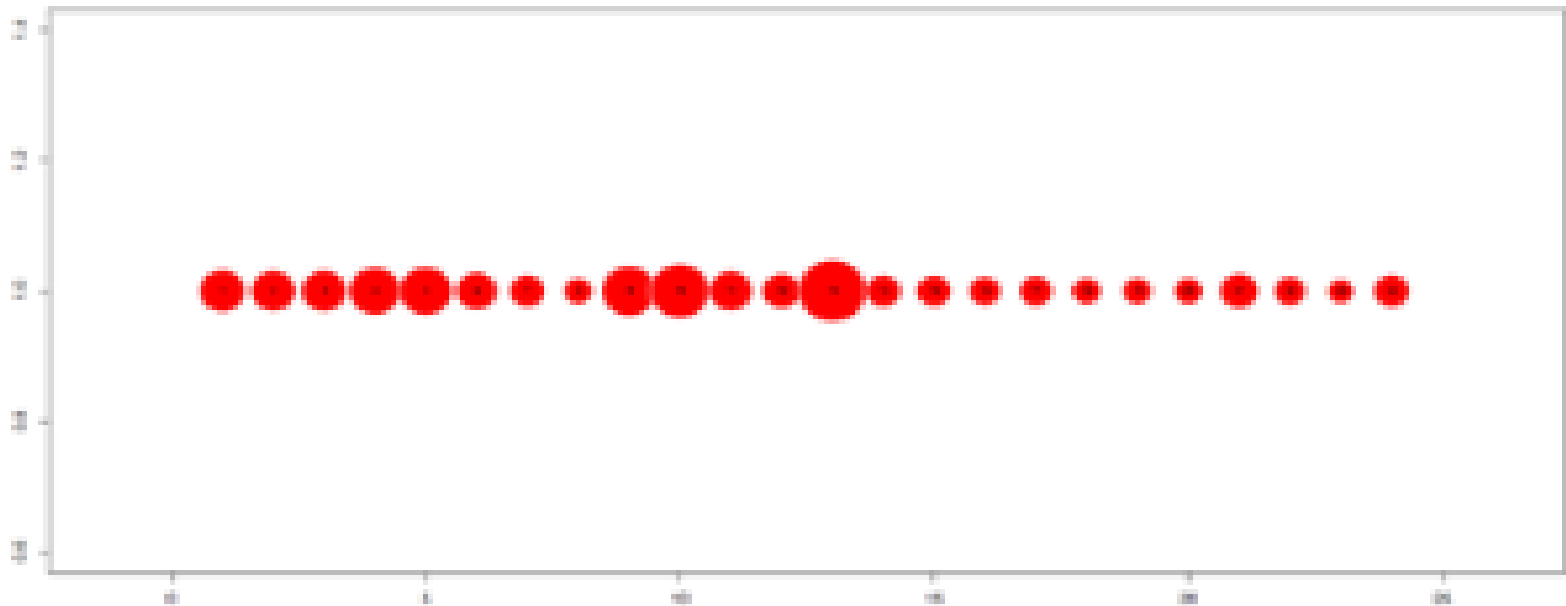
사례적용 추석명절 정체정보 시각화(1)

○ 추석명절 정체 정보 시각화

- 한국도로공사에서 보유한 빅데이터 중에 가장 요구사항이 높은 것은 명절 시기의 정체정보
- 명절에 따라서 데이터의 성격이 많이 틀린 것을 감안하여 추석으로 한정하여 정체정보 시각화를 진행
- 각 고속도로별 정체정보에 대한 복합적인 데이터 마이닝 작업이 필요 => 빅데이터를 다루는 전문가와의 협업이 필요
- 정보디자인측면의 요구사항을 정리하여 빅데이터 마이닝 전문가에게 전달하고 결과 데이터를 받는 과정에 많은 시간이 소요됨 => **효율을 위해서 정보디자인 담당자 역시 해당 빅데이터에 대한 이해와 이를 위한 데이터 마이닝 과정에 대한 이해가 필요함**

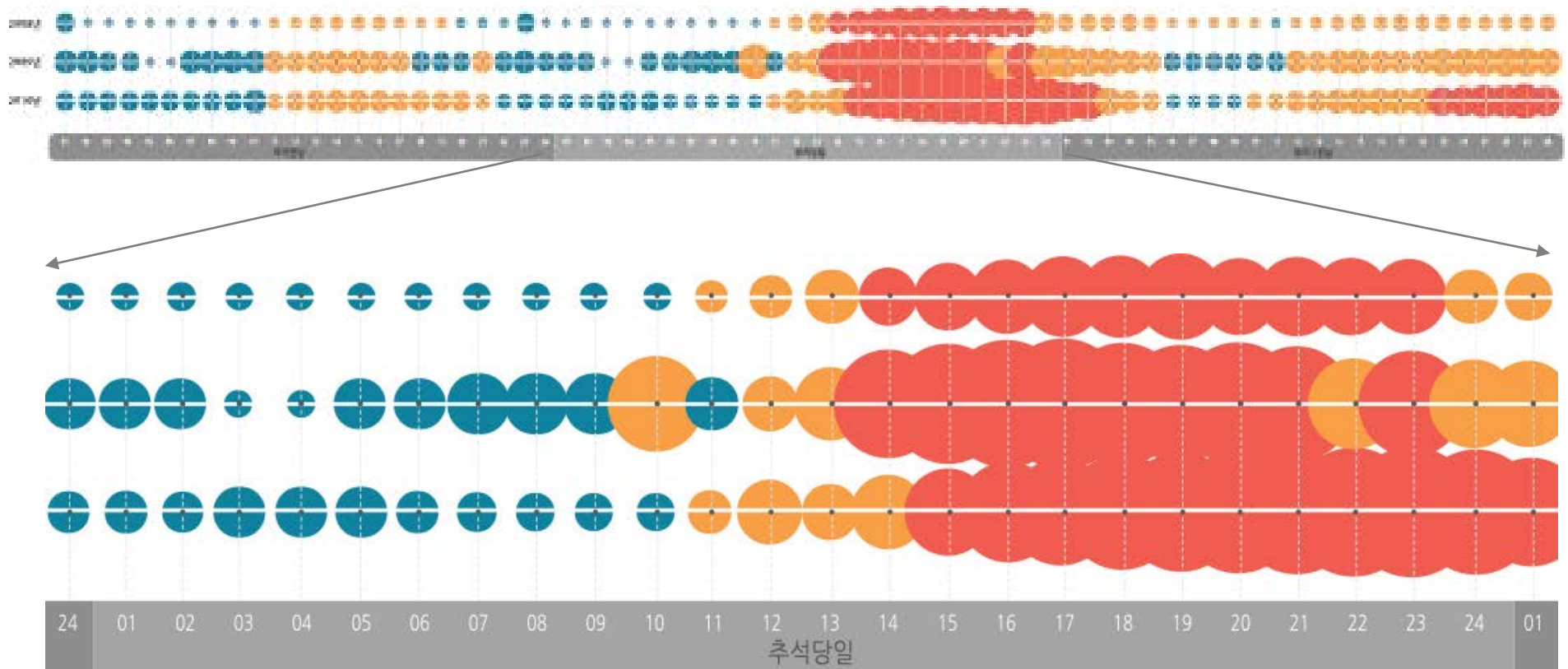
사례적용 추석명절 정체정보 시각화(3)

- 데이터 마이닝 작업 이후 R이라는 오픈 소스 통계 프로그램을 사용하여 데이터 그래프를 가공
- 서울을 기준으로 주요 5개 도시를 가는 데에 걸리는 시간을 표현하기에 가장 적합
- R을 사용하여 원의 '크기'를 이용한 버블차트의 형태의 벡터 추출



사례적용 추석명절 정체정보 시각화(4)

- 시각각 표현의 요소: **‘크기’, ‘색상’, ‘시간’ 다중 시각 표현기법 이용**
 - 1년 치를 한 줄에 3년 치를 세 줄에 표현
 - 세 가지의 색상과 원의 크기로 어느 시간이 전체적으로 막혔고, 어느 시간대가 원활했는지 직관적으로 시각화 개발



추석명절 고속도로 정체 시각화

단국 축제 명절도 일몰과 추석은 명절기간동안 대부분이 사람들이 고향을 돌아 떠난다. 일상에 갈 수 없, 온 가족이 있는 명절은 평소 중개은 고향가는 길이 바뀌어 추석은 고속도로 휴가객 시간 지체를 예상 할 수 밖에 없다. 2012년 추석은 연휴기간이 평일이와 다를 정도로 길다. 추석 연휴기간이 길수록 길의 길이 또한 고속도로 사용량과 밀접하게 관련이 있는 것으로 보인다. 고속도로 정체가 추석 명절에 비해 더욱 심해질 것으로 보인다.

가장 막히지 않는 시간 찾아보기

서울을 기준으로 주요 목적지(대전, 부산, 광주, 강릉, 목포)까지 고속도로 추석 연휴기간(대중이 휴가객) 소모시간을 차량의 정체 상황(정체)을 시각화 하여 보여주고 있다.

● 1차 ● 2차 ● 3차

서울 - 대전



서울 - 부산



서울 - 광주



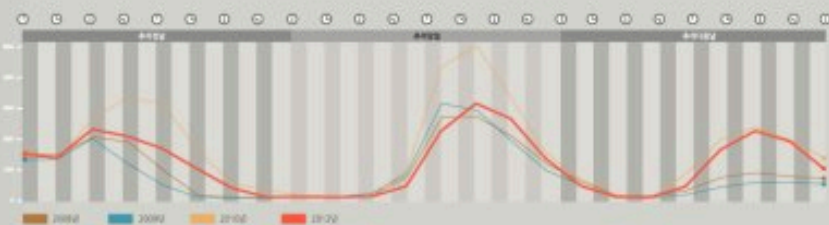
서울 - 강릉



서서울 - 목포



2012년도 추석연휴 예상 정체정도



추석명절에 얼마나 많은 차량이 이동하는가



과거 출발지에서 목적지까지 가장 많이 걸린 시간



결론...

결론 빅데이터 시각 표현

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

데이터 시각화 요건

- 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
+
인터랙티브 시각화
프로그래밍 능력

정보디자인의
시각표현
전문성

이지선

숙명여자대학교 시각영상디자인학과

jisunlee@sookmyung.ac.kr

facebook: jisunlee.net

twitter: jisunlee