

인터랙티브 데이터 시각화

WEEK

04

데잇걸즈

한국정보화진흥원

데이터와 인지 : 패턴

차트에 패턴 활용하기

송한나 Hannah Sookyong Song

2017.09.25.MON

A dark, grayscale background image showing a close-up of a spiral-bound notebook with a pen resting on it, and a portion of a computer keyboard visible in the upper left corner.

1

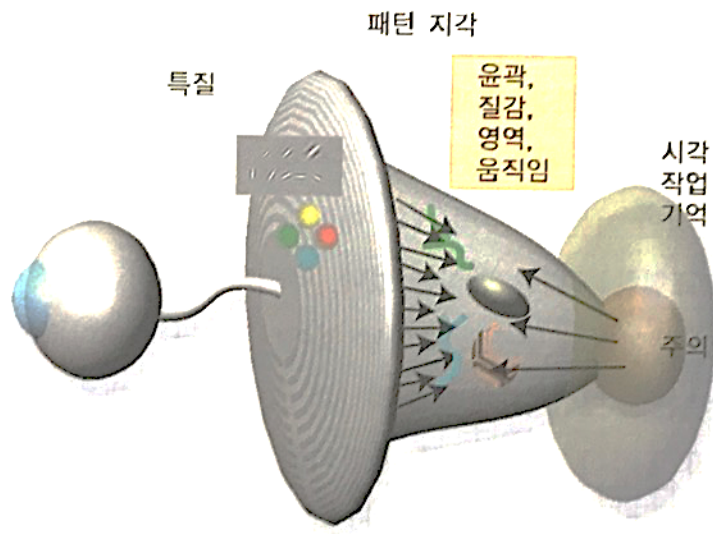
게슈탈트 법칙

게슈탈트 법칙

데이터 분석과 패턴

데이터 분석은 이전에는 알려져 있지 않았거나 표준에서 벗어난 **패턴을 찾는 것**이라고 할 수 있다. 예를 들어, 주식 시장 분석가는 수익의 변화를 예언하는 변수들의 패턴을 찾는다. 마케팅 분석가는 고객 데이터베이스에 있는 추세나 패턴을 인지한다. 과학자는 가설을 검증하거나 반박하는 패턴을 찾는다.

시각 뇌 visual brain는 **패턴을 찾는 강력한 엔진**이다. 데이터를 적절하게 시각적으로 변형시킬 수 있다면 데이터의 구조도 나타날 것이다. 패턴 지각은 대부분 2차원 공간에서 시각 장면의 특징 분석을 위한 2차원 처리 과정의 모음으로 볼 수 있다.



게슈탈트 법칙

게슈탈트 Gestalt : 독일어로 ‘패턴’을 의미한다.

게슈탈트 심리학자들은 많은 기초적 지각 현상을 명확하게 묘사했으며, 패턴 지각의 게슈탈트 법칙을 만들어냈다.

게슈탈트 법칙

근접성 proximity

유사성 similarity

연결성 connectedness

연속성 continuity

대칭성 symmetry

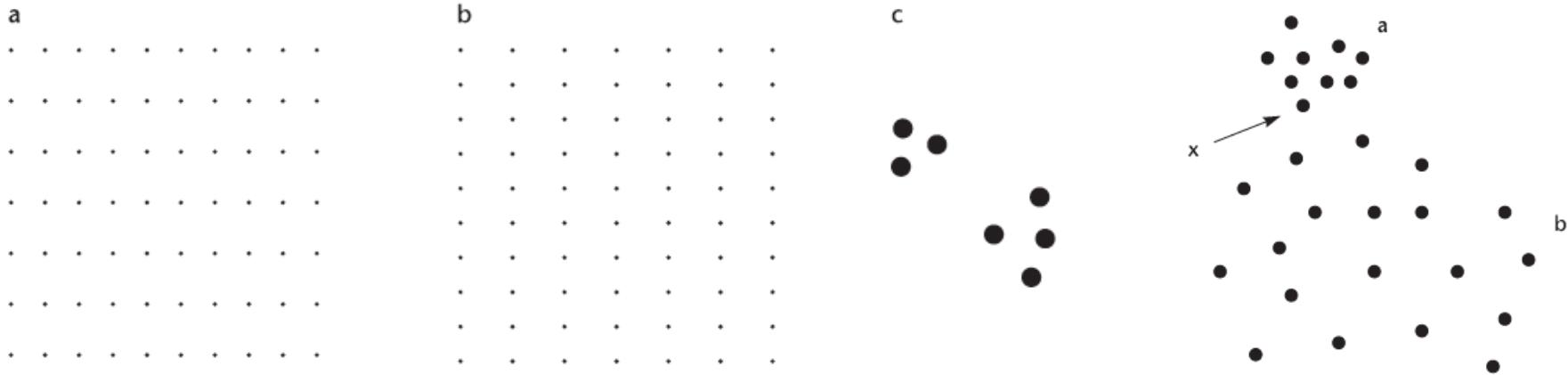
폐쇄성 closure

상대적 크기 relative size

공동 운명 common fate

근접성 proximity

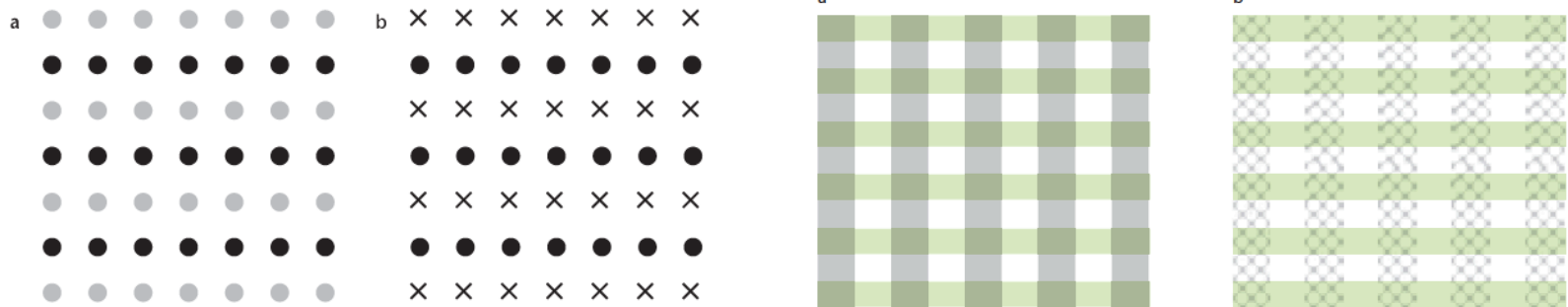
근접성 : 가까이 있는 물체들은 지각적으로 함께 그룹화된다.



→ 서로 관련 있는 정보는 기호와 글리프를 서로 가까이 배치한다.

유사성 similarity

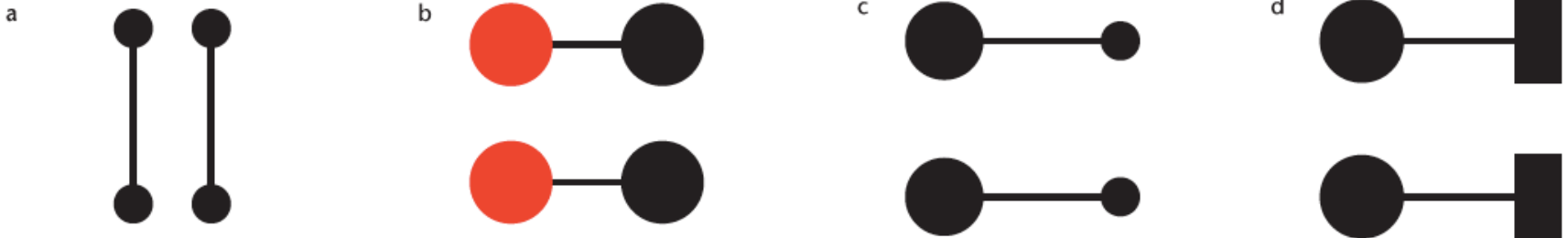
유사성 : 유사한 요소들은 같이 그룹화되는 경향이 있다.



→ 데이터 세트의 격자 grid 배치를 디자인할 때 색과 질감 같은 저수준의 시각 특성을 사용해 행과 열, 혹은 행이나 열로 표현하도록 한다.

연결성 connectedness

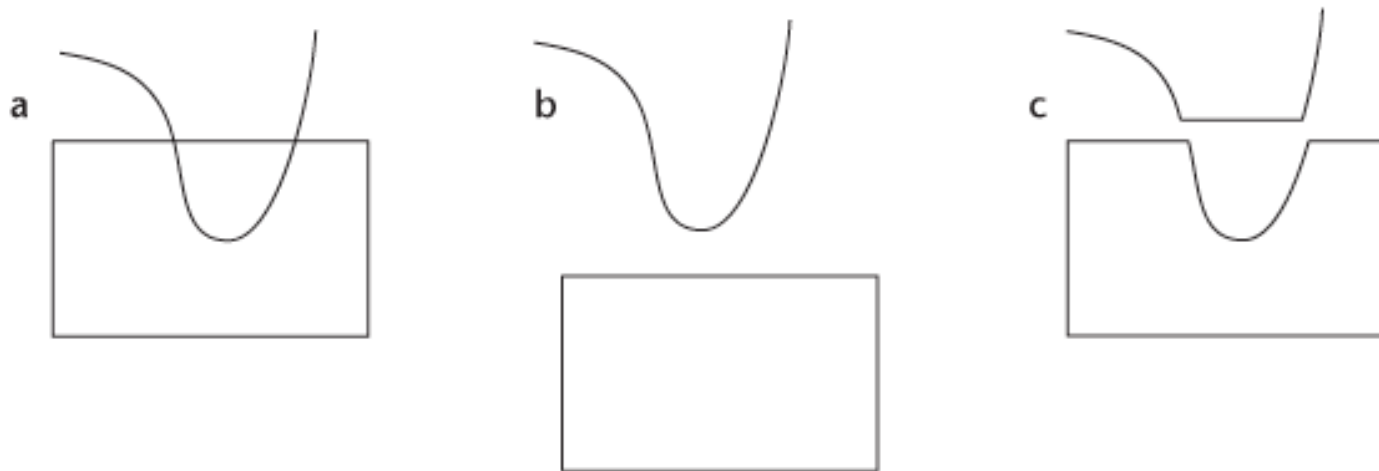
연결성 : 근접성보다 더 강력한 그룹화 원리가 될 수 있다. 각기 다른 그래픽 대상들을 선으로 연결하는 것은 대상들 간에 어떤 관계가 있다는 것을 나타내는 매우 강력한 방식이다. 이것은 개념들 간의 관계를 보여주는 가장 흔한 방식 중 하나인 노드 링크 다이어그램의 근본이 된다.



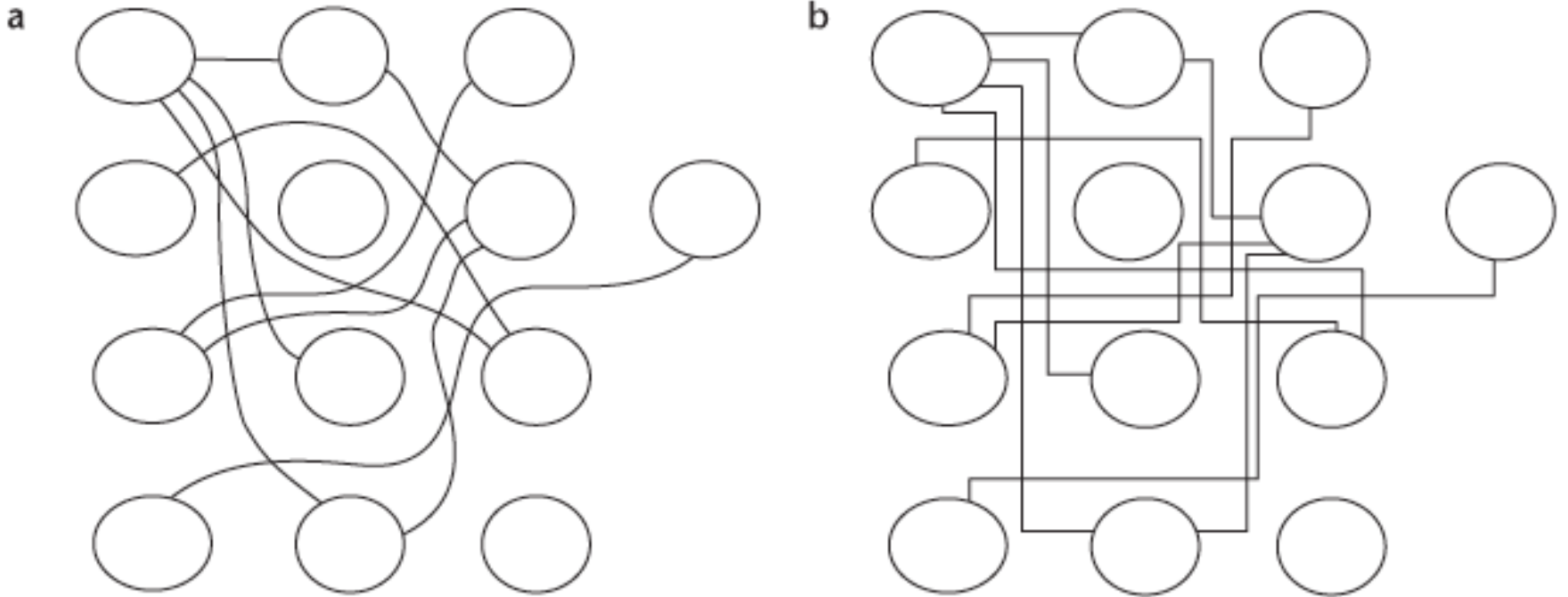
→ 객체 사이의 관계를 보여주려면 선이나 색 리본을 이용해 데이터의 그래픽 표현들을 연결시키도록 한다.

연속성 continuity

연속성 : 갑작스런 방향 변화가 있는 시각적 요소들보다 완만하고 지속적인 시각 요소들로부터 시각적 객체들을 구성하게 된다. 바람직한 연속성의 원리는 노드들과 연결선으로 구성된 다이어그램을 그릴 때 적용할 수 있다. 연결선들의 시작점과 도착점이 완만하고 지속적인 경우에 더욱 확인하기 쉽다.



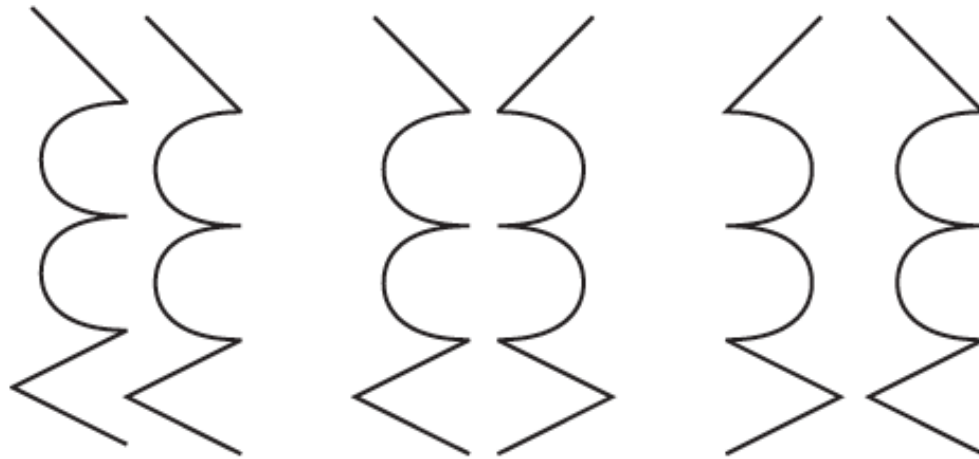
연속성 continuity



부드러운 곡선으로 연결된 노드 다이어그램이 훨씬 더 지각하기 쉽다.

대칭성 symmetry

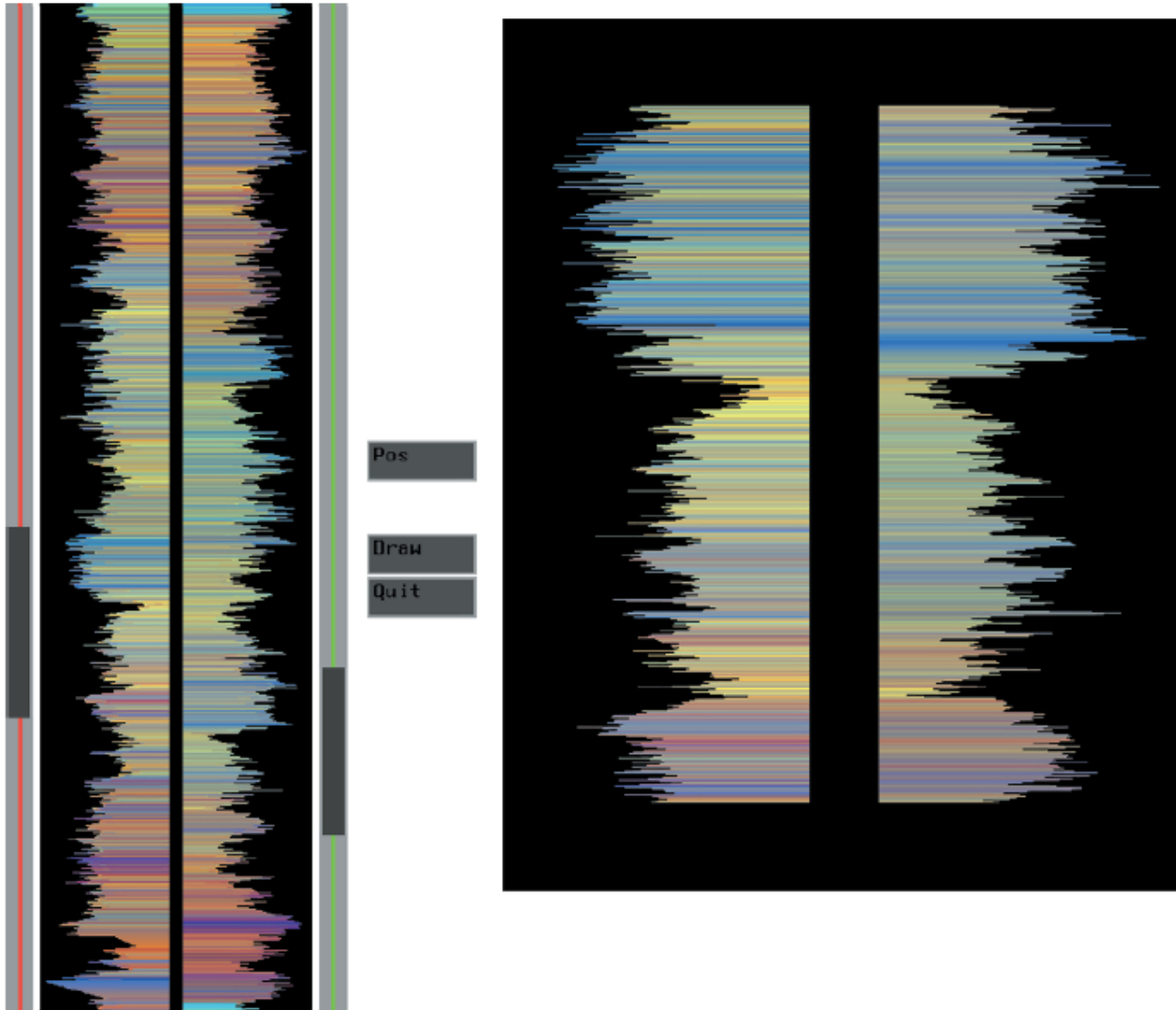
대칭성 : 대칭적으로 배치된 선의 쌍은 평행선의 쌍보다는 더 강력하게 시각 전체를 형성하는 것으로 지각된다. 대칭은 데이터 분석가들이 두 개의 다른 시계열 데이터 간의 유사성을 찾는 과제에 응용할 수 있다.



동일한 평행 윤곽선으로 구성되어있으나,
수직축에 대칭인 패턴이 훨씬 강력하게 전체적인 형태로 인지된다.

→ 패턴을 더 쉽게 비교하려면 대칭을 사용한다. 비교되는 패턴은 시각도 면에서 작아야(가로 1도 이하, 세로 2도 이하) 하며, 가로 혹은 세로로 배열한다.

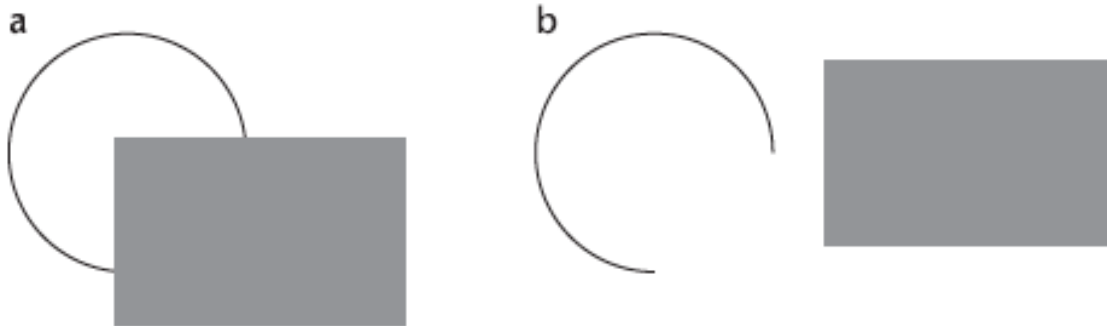
대칭성 symmetry



폐쇄성 closure

폐쇄성 : 폐쇄된 윤곽선은 물체로 보이는 경향이 있다. 게슈탈트 심리학자들은 안에 틈이 있는 윤곽선들을 닫아버리려는 지각적 성향이 있다고 주장했다.

어떤 폐쇄 윤곽선이든 윤곽선 공간을 윤곽선의 '내부' 혹은 '외부'로 구분하려는 강한 지각적 성향이 있다. 윤곽선으로 둘러싸인 영역은 공통 영역(common region)이 된다. 공통 영역이 단순한 근접성보다 더 강력한 조직화 원리이다.

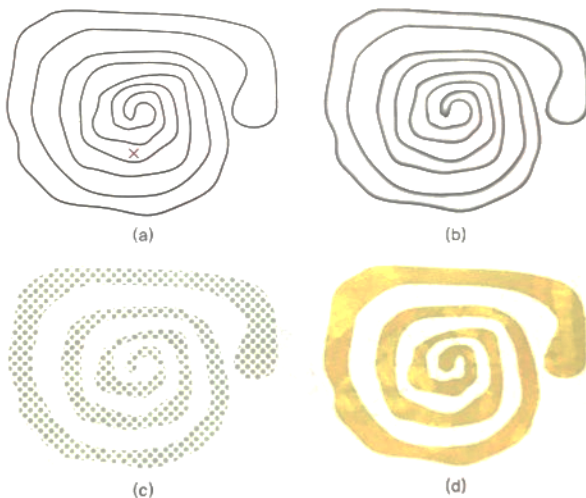


절단된 원이 아니라 직사각형 뒤에 있는 원을 생각하게 된다.

폐쇄성 closure

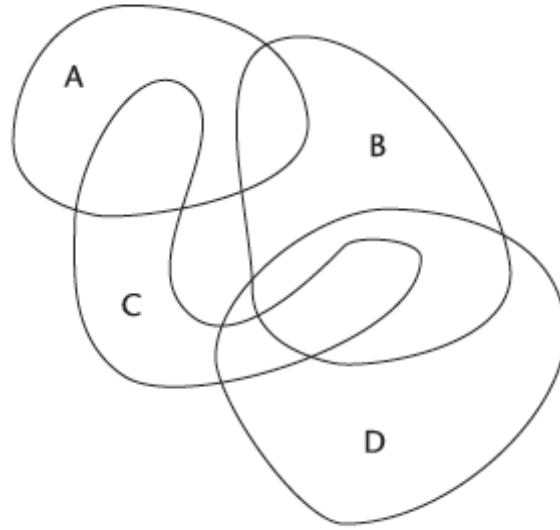
폐쇄 윤곽선은 벤 오일러 Venn-Euler 다이어그램의 집합 개념을 시각화하는 데 널리 사용된다. 오일러 다이어그램에서 폐쇄 윤곽선의 내부 영역을 요소의 집합을 정의하는 것으로 해석한다. 다수의 폐쇄 윤곽선은 각기 다른 집합들 간의 중첩되는 관계를 나타내기 위해 사용된다. 벤 다이어그램은 중첩이 가능한 모든 영역을 포함하고 있는 오일러 다이어그램의 제한된 형태이다.

이러한 종류의 다이어그램에서 가장 중요한 시각적 요소 두 가지는 폐쇄성 closure와 연속성continuity이다.



영역이 복잡할 때 단순한 윤곽선은 부적합하다.
예지, 질감, 색을 이용해 더 확실하게 규정될 수 있다.

폐쇄성 closure



오일러 다이어그램

a



b



질감과 색을 이용해 확장된 오일러 다이어그램은
폐쇄 윤곽선만을 사용한 기존의 오일러 다이어그램보다 더 복잡한 관계의 집합을 표현할 수 있다.

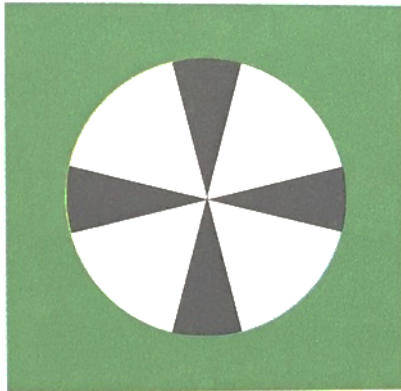
폐쇄성 closure



호텔(주황색), 지하철 역(갈색), 의료 기관(보라색)의 분포를 명시하기 위해 윤곽선으로 규정된 영역과 색으로 규정된 영역을 함께 추가했다.

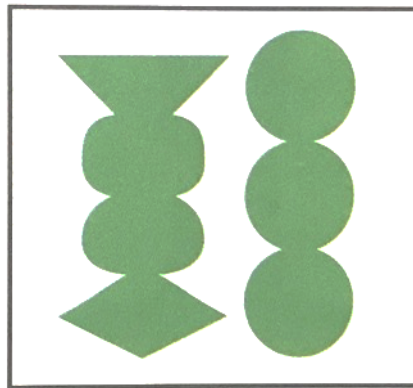
전경-배경 figure-ground

전경-배경 figure-ground 효과 : 전경이란 앞에서 지각되는 물체와 같은 것이다. 배경이란 그 전경 뒤에 놓여있는 것을 말한다. 일반적으로 어떤 패턴의 작은 구성 요소들은 물체로 지각되는 경향이 있다.



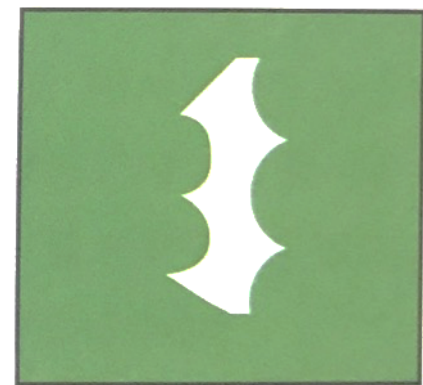
(a)

검정색 영역이 더 작기 때문에 물체로 지각되기 더 쉽다. 또한 수평, 수직 방향의 패턴도 물체로 지각하는 것이 더 쉽다.



(b)

녹색 영역이 전경으로 보인다.



(c)

녹색 영역은 전경으로 보이지 않는다.

전경-배경 figure-ground



꽃병의 지각은 대칭과 폐쇄 영역에 의해, 얼굴로 지각되는 것은 사전 지식에 의해 이뤄진다. 고수준과 저수준 처리 간 경쟁이 발생한다.

2

벡터의 표현

방향과 방위

벡터장 표현 : 방향과 방위 지각

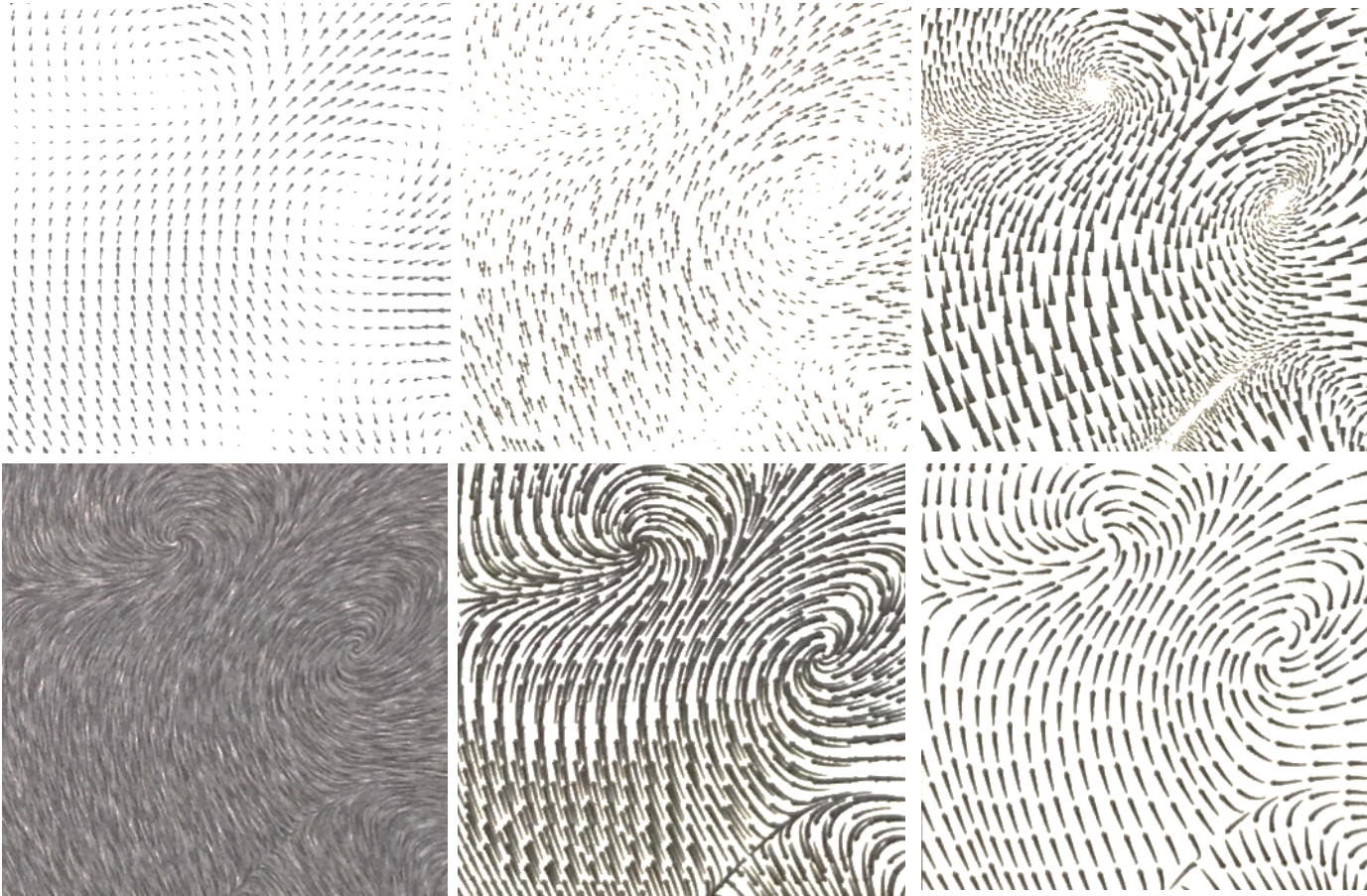
벡터 vector를 표현하는 데 발생하는 문제점은 세 가지 요소로 구분되는데, 벡터 규모의 표현, 방위 표현, 특정 방위에서의 방향 표현이다. 어떤 기법들은 한 개나 두 개의 요소를 표현하지만 세 개를 표현할 수는 없다. 예를 들어 바람의 속도(규모)를 색을 이용해 스칼라장으로 표현할 수 있다.



벡터장 vector field 데이터의 표현에는 윤곽선 지각 이론을 활용할 수 있다. 방향성이 있는 요소들의 규칙적인 격자를 만들어낸다. 시작에서 끝으로 향하는 배열이 흐름의 패턴을 보기 쉽게 만든다.



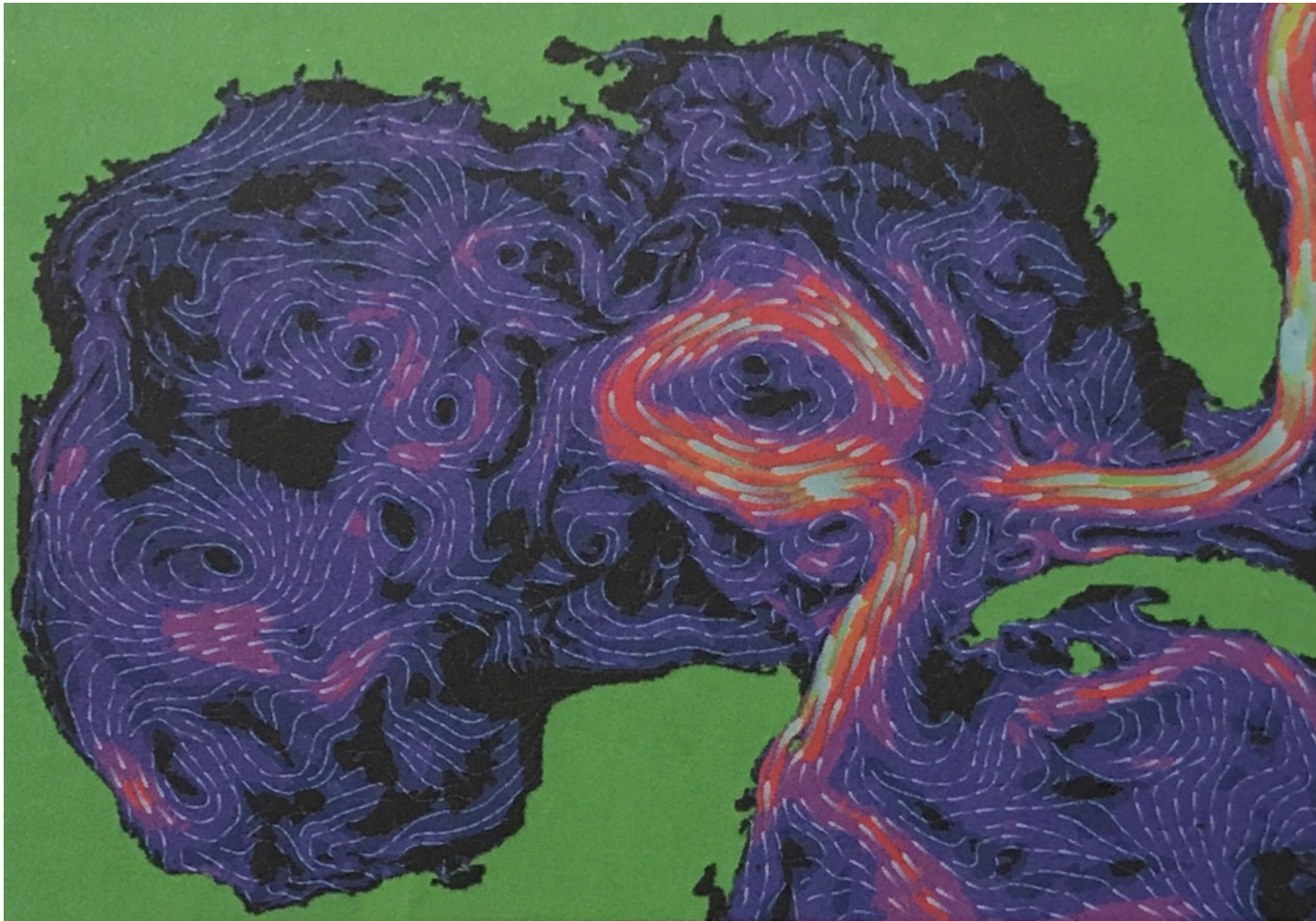
벡터장 표현 : 2차원 흐름의 시각화 기법



- (1) 규칙적인 격자무늬 위의 화살표
- (2) 지각적 알리아싱 효과를 줄이기 위한 흐트러진 격자무늬 위의 화살표
- (3) 아이콘의 크기가 장의 세가와 비례하고 밀도와 반비례하는 삼각형 아이콘

- (4) 선 적분 정리
- (5) 화살표가 흐름선을 따라 존재하며 규칙적인 격자를 사용
- (6) 화살표가 흐름선을 따라 존재하며 일정한 간격 알고리즘을 사용

벡터장 표현



멕시코만 표면 해류. 머리-꼬리 요소들이 사용됐으며, 각 요소는 머리 부분이 꼬리보다 더 뚜렷하다. 속도는 폭, 길이, 배경색으로 표현된다.

3

다차원 이산 데이터 의 패턴 지각

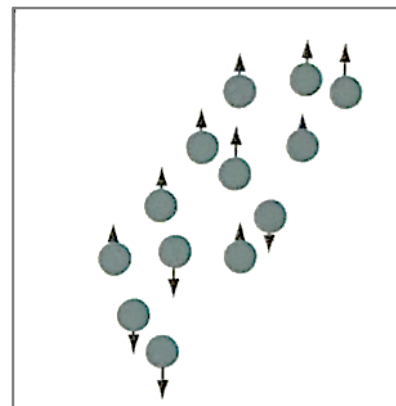
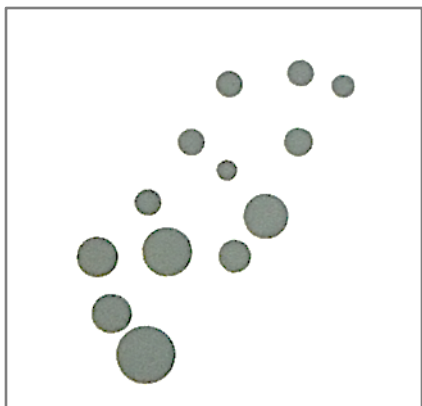
다차원 이산 데이터의 패턴 지각

다차원 데이터의 시각화

데이터 탐색을 위한 방법으로, 플로팅plotting은 데이터를 시각화하여 패턴을 찾고 발견한 것을 해석하는 과정이다. 패턴에는 군집 cluster와 상관관계 correlation의 유형이 있다.

색상과 크기의 추가

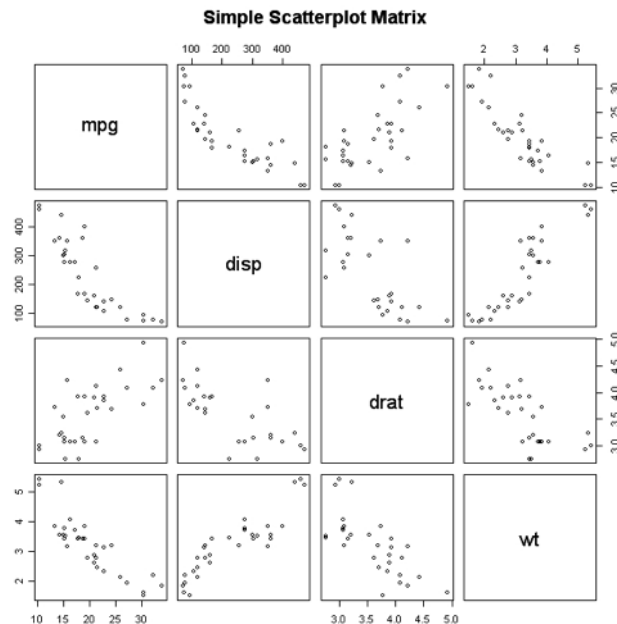
두 개 이상의 수적 속성이 포함되는 경우 보통 글리프 크기와 색을 추가한다. 크기, 음영, 변동하는 점의 움직임 위상 중에서 위상 차이에 가장 민감하다. 그러나 색이나 크기의 사용도 3차원이나 4차원 데이터를 나타내는 방법으로 활용된다.



다차원 데이터의 시각화

산점도 행렬 scatterplot matrix

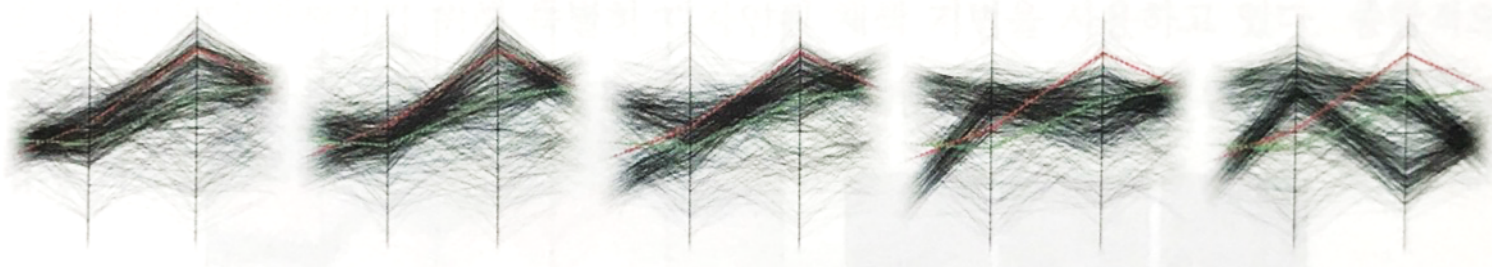
3차원 이상의 고차원 데이터의 표현은 산점도 행렬 scatterplot matrix로 해결할 수 있다. 이 기법에서 모든 변수의 쌍은 이차원 산점도를 만들어내는 데 사용된다. 일반화된 산점도 행렬은 종종 유용하게 사용되지만, 3차원 또는 더 높은 차원이 동시에 고려될 때만 이해될 수 있는 고차원 데이터 패턴은 보기 어렵다.



다차원 데이터의 시각화

평행좌표계 parallel coordinates plot

두번째 방법은 평행좌표계 parallel coordinates plot이다. 이 그래프에서 각 속성 차원은 수직선으로 표현되고, 데이터 지점들은 다양한 속성 값들을 연결하는 선이 된다. 이 기법은 지각되는 패턴들은 축이 서로에 대해 정렬되는 방식에 의해 결정된다는 것이다. 축의 순서를 변경하면 지각되는 패턴들도 바뀐다.



그러나 평행좌표계는 브러싱 brushing이라는 상호작용 기법과 함께 사용된다. 브러싱으로 선택한 축에서 범위가 선택되며 그 범위를 통과하는 폴리라인이 강조되게 된다. 이 방법은 데이터 탐색 과정의 일부가 된다.

Q&A

e-mail : cognitumlab@gmail.com

web : www.cognitumlab.com