

Basic of data visualization



데이터의 차원: 1차원

- 하나의 차원 층위를 가진 데이터 의미
- 기울기 없는 직선 그래프, 막대 그래프, 파이 도표 등

	2호선		
역 이름	거리(km)	누적 거리	시간
시청	0.0	0.0	
을지로 입구	0.7	0.7	2
을지로 3가	0.8	1.5	2
을지로 4가	0.6	2.1	1
동대문역사문화공 원	1.0	3.1	2
신당	0.9	4.0	2
상왕십리	0.9	4.9	2
	•••	•••	•••

역 이름_순차 항목 [n번째 역]=[구간 거리]



🔎 데이터의 차원: 1차원 + 시간

- 어제, 한달 전, 1년 전 특별한 주기를 이용하여 데이터를 측정
- 하지만 데이터가 동일하게 측정되었다면 이는 정보라 부를 수 없음
- 시간에 따른 변화를 기록한 그래프 > 주식 시장 그래프, 대통령 선거의 사전조사 변화 추이 분기별 물가 상승 곡선 등...
- 시간과 함께 변화하는 데이터는 우리 일상 그 자체이며, "시계열 데이터"라 부름

시계열 데이터는 "값", "시간" 이라는 두 개의 축을 가지고 있으므로 2차원 데이터라고 볼 수 있음

> 시간이라는 축은 언제나 같은 속도로 어느 누구에게나 혹은 어떤 사물에나 공평하게 적용되는 예외가 없는 축



데이터의 차원: 1차원 + 시간

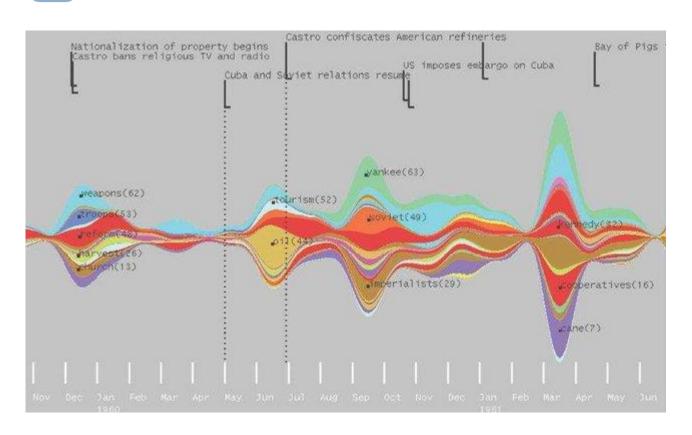
- 정보 이론의 엔트로피 (entropy) : 무작위성과 유사한 뜻
- 엔트로피가 높다 = 예측 불가능하다 = 정보가 많다
- 규칙성이 있거나 노이즈를 제거한 데이터에는 정보가 적다 라고 말함
- 그러므로 정보가 많고 = 엔트로피가 높고 = 예측 불가능성이 크면
 > 압축하기가 힘들고 암호를 깨기도 힘들다

시간.. 정량적으로 증가,, 변화가 없는 것과 마찬가지 = 정보가 없다 라고 할 수 있음

즉, 엄밀히 말해 시간 데이터는 정보를 주는 것이 아니므로 정보성이 떨어진다라고 정리할 수 있음

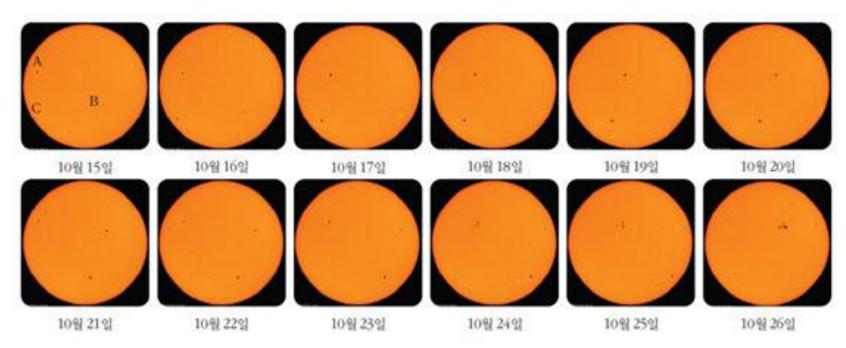
P

데이터의 차원: 1차원 + 시간



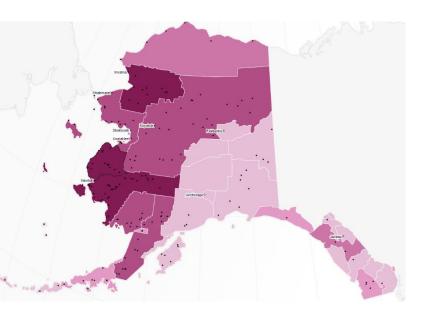


데이터의 차원: 1차원 + 시간



🔎 데이터의 차원: 2차원

- 서로 다른(독립된) 두 가지 차원을 가지고 있음예) 위, 경도라는 지구 위치 정보 데이터
- 2차원 시각화 원조 지도 (카토그라피, cartography)
- 우리나라 카토그라피 대동여지도



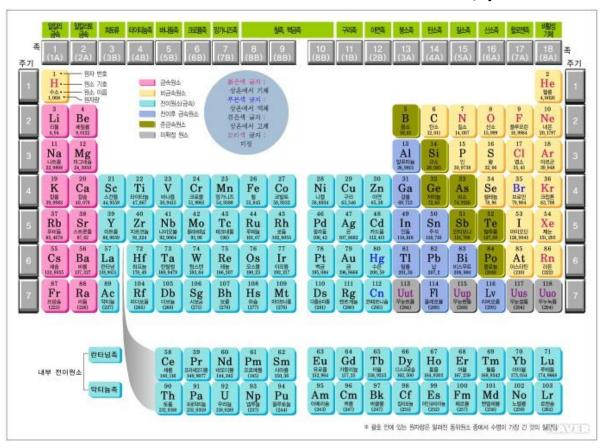






데이터의 차원: 2차원

- 전체적인 데이터를 1차원으로 해석하는 것이 자연스러운 상태에서 이를 2차원 배열로 바꾸었을 때, (원소 주기율표)





데이터의 차원: 다차원

- 실제 세상은 훨씬 복잡
- 실제 대부분 데이터 집합은 세 개 이상 속성으로 구성
- 다차원(또는 다변수) 데이터 사례 > 관계 데이터베이스 테이블
- 표현 했지만 읽는 법을 모르면 마찬가지로 혼란스러움 (해당 분야에 대한 기본 지식)
- 네트워크, 상호 참조 그래프, 산포도 (3차원) 등 형태

