데이터 시각화

발표내용

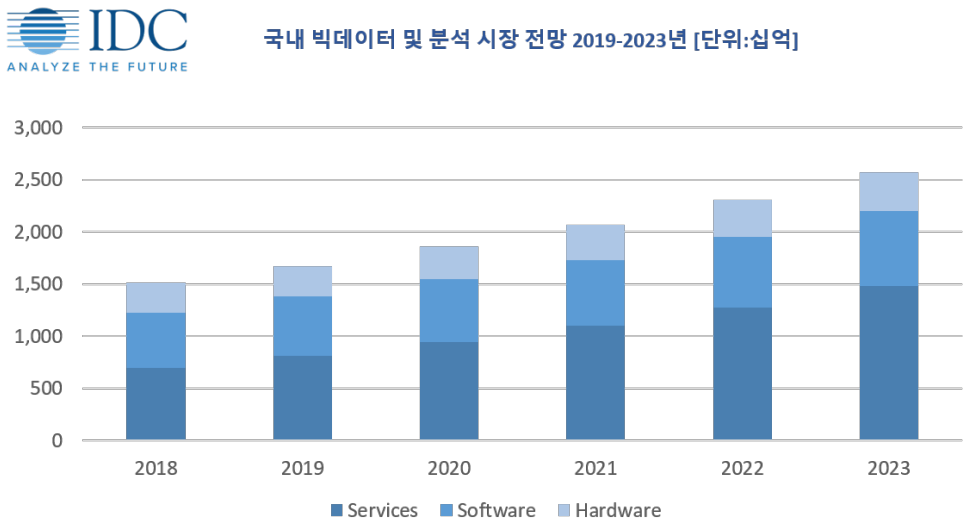
- 시각화 정의 및 이해
- 시각화의 방법
- 시각화 사례 소개

데이터 시대의 도래

2019년 빅데이터 시장 : **1조 6744억원**

21 Jan 2020

한국IDC, 국내 빅데이터 및 분석 시장 연평균 11.2% 성장해 2023년 2조 5,692억원 규모 전망



Source: IDC, 2019

2019년 데이터 직무 인력을 분야별

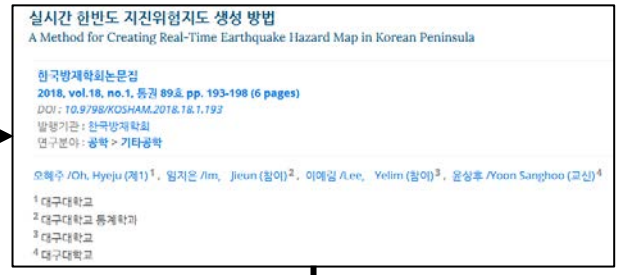
- 데이터 개발자(35.2%)
- 데이터 엔지니어(16.4%)
- 데이터베이스 관리자(14.5%)
- 데이터 분석가(7.4%)
- 데이터 아키텍트(7.3%)
- 데이터 컨설턴트(6.4%)
- 데이터 과학자(1.8%, 1600여 명)

출처 : 미래한국 Weekly(http://www.futurekorea.co.kr)

사례로 보는 데이터시각화

번호	발생시각	규모	깊이 (km)	최대 진도	위도	경도	위치	지도보기	상세정보 (진도 등)
53	2020/10/03 01:37:30	2.3	-	I	38.83 N	126.03 E	북한 황해북도 연산 서남서쪽 20km 지역	지도	상세정보
52	2020/10/02 22:29:50	2.6	-	I	38.74 N	125.90 E	북한 황해북도 송림 동쪽 23km 지역	지도	상세정보
51	2020/10/02 21:31:23	2.5	-	I	38.74 N	125.89 E	북한 황해북도 송림 동쪽 22km 지역	지도	상세정보
50	2020/09/22 20:08:54	2.1	15	II	36.60 N	129.37 E	경북 영덕군 북쪽 21km 지역	지도	상세정보
49	2020/09/20 01:22:14	2.2	17	III	36.51 N	127.14 E	충남 공주시 북북동쪽 7km 지역	지도	상세정보
48	2020/09/13 00:54:03	2.9	15	I	37.64 N	124.60 E	인천 옹진군 백령도 남남서쪽 35km 해역	지도	상세정보
47	2020/09/12 13:31:05	2.4	15	II	36.03 N	128.40 E	경북 칠곡군 북쪽 4km 지역	지도	상세정보
46	2020/09/02 19:43:06	2.3	-	II	34.38 N	126.20 E	전남 진도군 남남서쪽 13km 지역	지도	상세정보
45	2020/08/21 03:52:51	2.5	14	II	37.91 N	125.93 E	북한 황해남도 해주 남동쪽 24km 지역	지도	상세정보
44	2020/08/07 21:27:43	2.4	-	I	41.30 N	129.09 E	북한 함경북도 길주 북북서쪽 43km 지역	지도	상세정보
43	2020/08/05 01:22:48	2.0	7	I	38.85 N	125.71 E	북한 황해북도 송림 북북동쪽 14km 지역	지도	상세정보
42	2020/08/01 13:15:06	2.0	19	II	36.77 N	128.48 E	경북 예천군 북쪽 13km 지역	지도	상세정보
41	2020/08/01 02:58:47	2.0	17	I	38.65 N	127.16 E	북한 강원 평강 북북서쪽 29km 지역	지도	상세정보
40	2020/07/25 00:34:06	2.2	7	I	34.06 N	125.54 E	전남 신안군 흑산도 남쪽 66km 해역	지도	상세정보
39	2020/07/20 19:02:00	2.0	11	I	34.53 N	128.14 E	경남 남해군 남남동쪽 41km 해역	지도	상세정보
38	2020/07/15 15:20:47	2.4	-	I	34.88 N	124.68 E	전남 신안군 흑산도 서북서쪽 77km 해역	지도	상세정보
37	2020/07/11 22:38:55	2.7	-	I	37.88 N	129.58 E	강원 동해시 북동쪽 57km 해역	지도	상세정보
36	2020/07/07 14:08:23	3.3	12	III	37.19 N	125.55 E	인천 옹진군 연평도 남남서쪽 54km 해역	지도	상세정보
35	2020/07/03 13:59:31	2.3	6	III	37.18 N	128.71 E	강원 영월군 동쪽 22km 지역	지도	상세정보

https://www.weather.go.kr/weather/earthquake_volcano/domesticlist.jsp



실시간 한번도 지진위험지도 생성 방법

A Method for Creating Real-Time Earthquake Hazard Map in Korean Peninsula

한국방재학회논문집

2018, vol.18, no.1, 통권 89호, pp. 193-198 (6 pages)

DOI : 10.9798/KOSHAM.2018.18.1.193

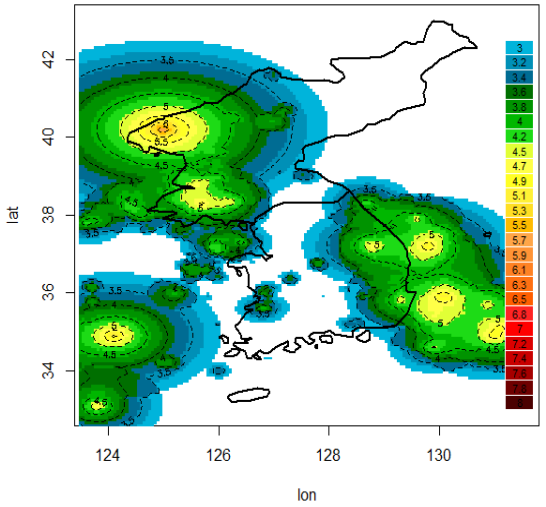
발행기관 : 한국방재학회

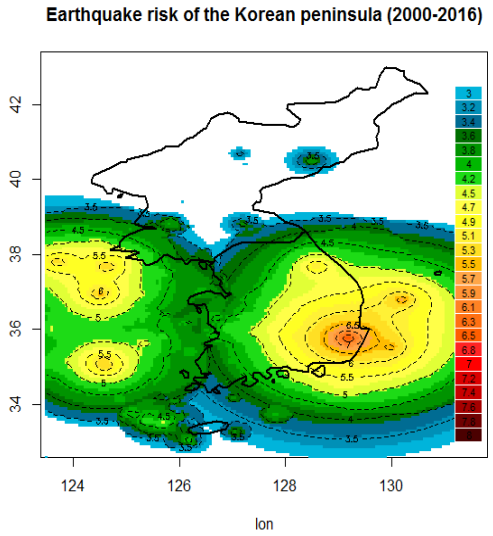
연구분야 : 공학 > 기타공학

오혜주 /Oh, Hyeju (제1)[1], 임지은 /Im, Jieun (참여)[2], 이예림 /Lee, Yelim (참여)[3], 윤상후 /Yoon Sanghoo (교신)[4]

[1] 대구대학교
[2] 대구대학교 통계학과
[3] 대구대학교
[4] 대구대학교

사용 프로그램 : R

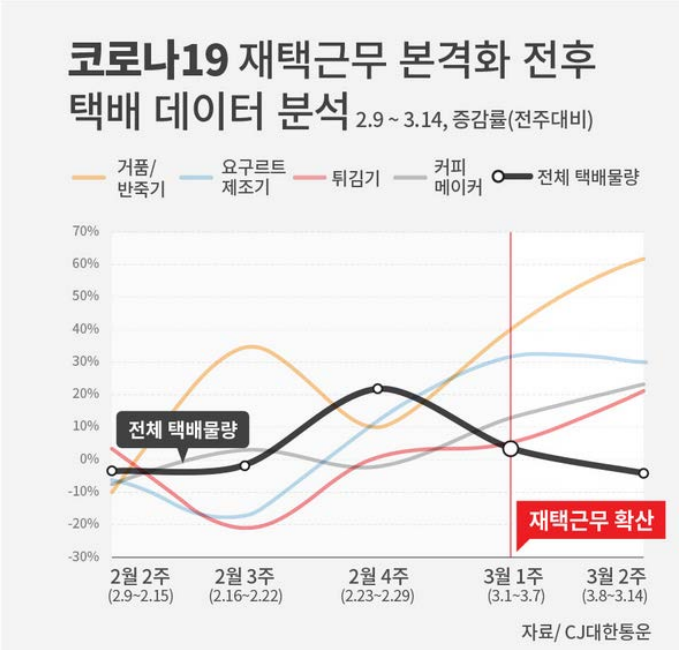




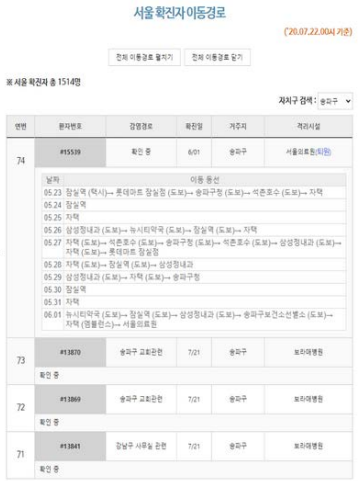
사례로 보는 데이터시각화

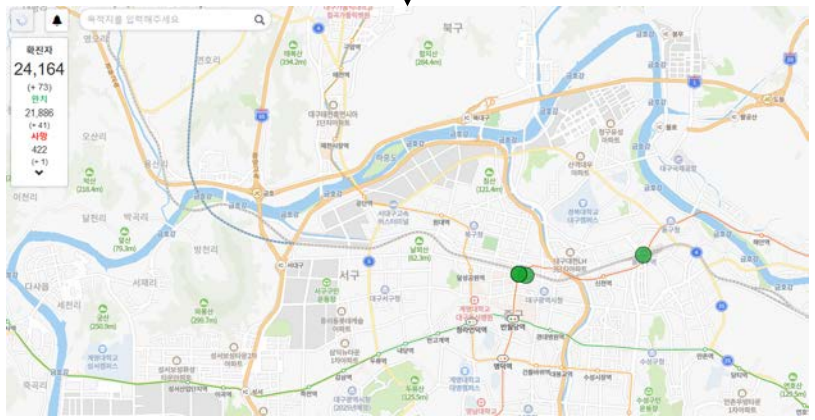
데이터 출처 : 2015년 전국 고액체납자 명단

사례로 보는 데이터시각화



감염 정보 수집





생필품 사재기에 대한 불안과 논란에 대해 총 1억 8천 건 배송의 시각화를 통해
사재기로 보이는 주문은 3일에 불과해 택배가 위기 해결에 기여 했다고 홍보

콘텐츠로서의 데이터

무엇을 느끼셨고

어떤 생각이 들었나요?

읽는 데이터에서 보는 데이터로

앞서봤던 사례처럼 데이터는

그저 **쌓아놓으려고** 저장하는 것은 아님

전통산업에선 데이터(정보)는 서비스 대상이 아님 = **엑셀**

3차 산업이 발달되면서 데이터는 콘텐츠로 진화 = **시각화**

기록 = 데이터

약 100년 전의 데이터

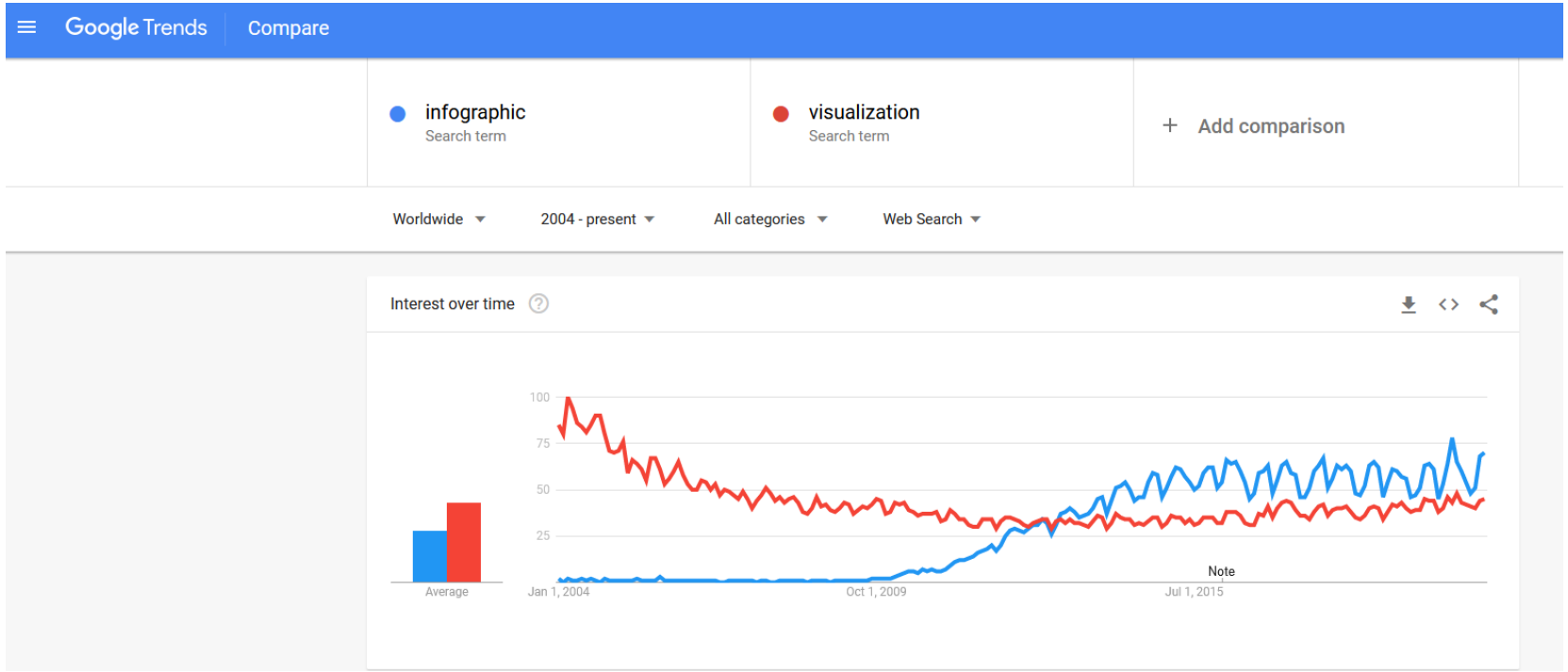
쌓아 놓고만 있는 데이터

불과 얼마 전까지의 'DB'

GOOGLE 서비스 (트렌드)

- 구글 트렌드를 통한 검색 관심도 및 점유율

https://trends.google.co.kr/trends/?geo=KR



한국인이 가장 많이 이용하는 검색 사이트?

NAVER

네이버를 이용하여 키워드를 알아보자.

NAVER DataLab.

데이터랩 홈 | 급상승검색어 | 검색어트렌드 | 쇼핑인사이트 | 지역통계 | 댓글통계

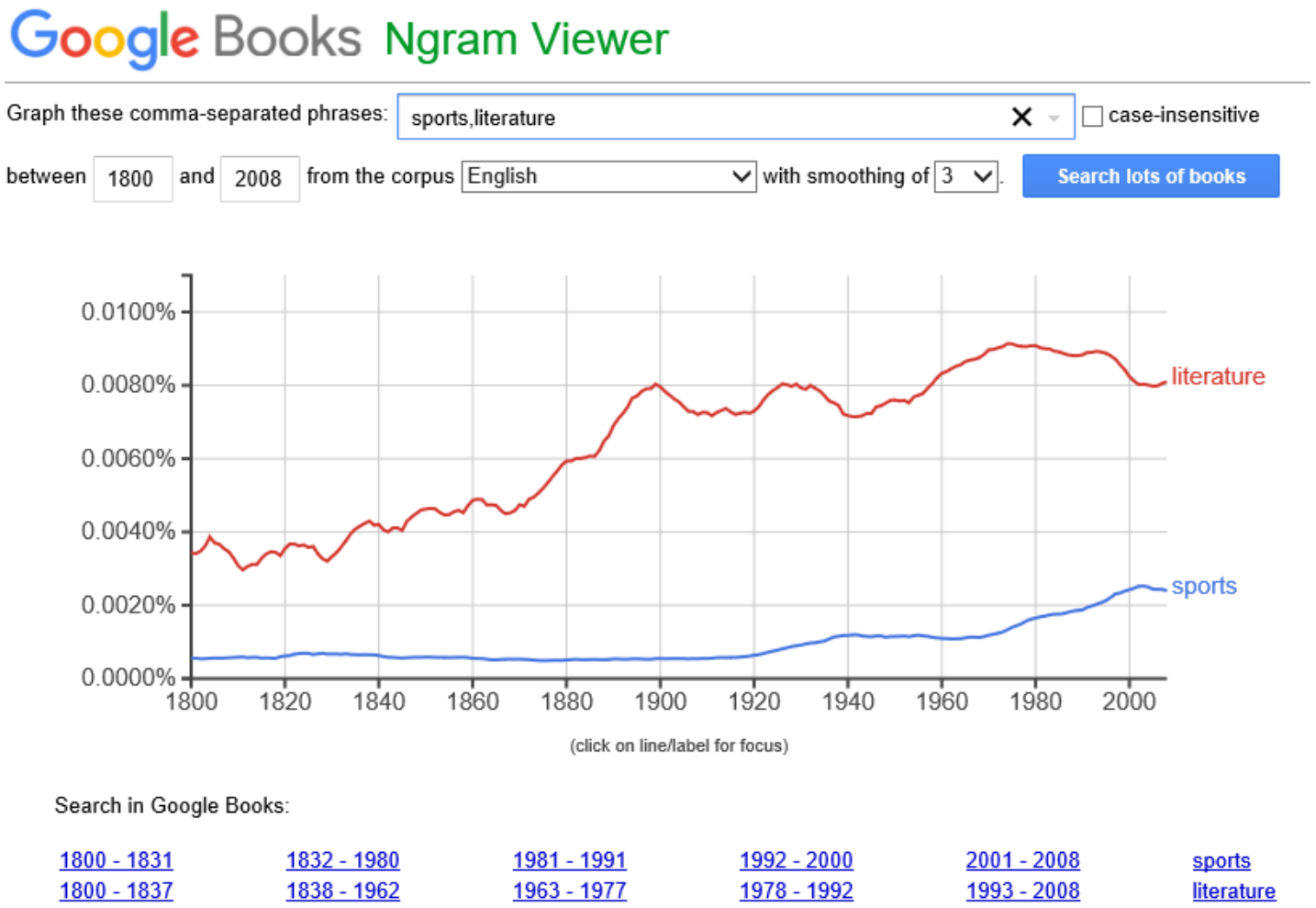
쇼핑인사이트 쇼핑 분야별 클릭 추이와 분야별 검색어 현황을 확인할 수 있습니다.

분야별 인기 검색어 | 인기분야

패션의류 | 일간

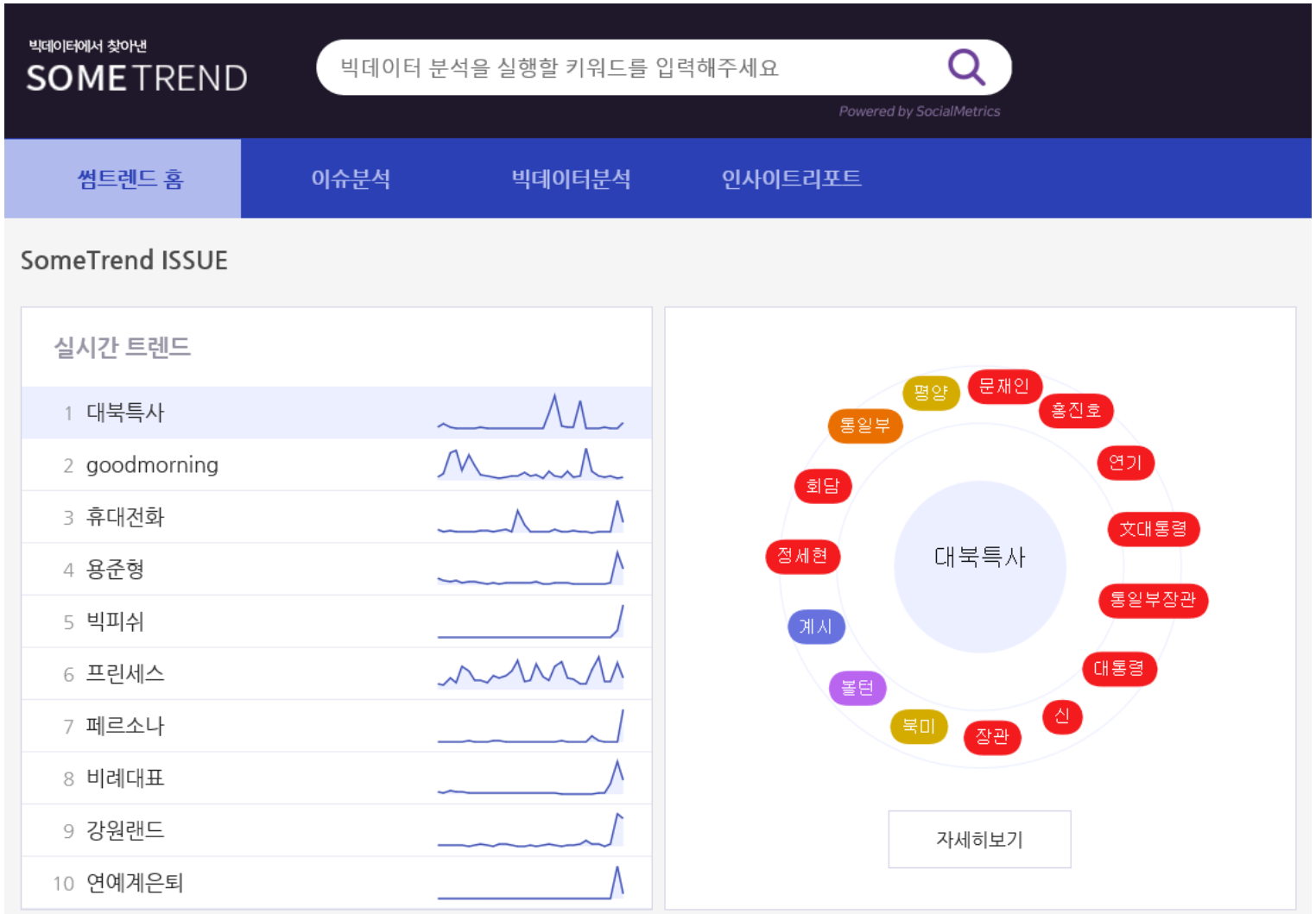
2020.10.01.(목)	2020.10.02.(금)	2020.10.03.(토)	2020.10.04.(일)
1 해리슨캐시미어코트	1 해리슨캐시미어코트	1 해리슨캐시미어코트	1 원피스
2 아디다스	2 원피스	2 원피스	2 나이키맨투맨
3 원피스	3 나이키맨투맨	3 나이키맨투맨	3 맨투맨
4 나이키	4 나이키후드티	4 맨투맨	4 나이키후드티
5 나이키맨투맨	5 맨투맨	5 나이키후드티	5 나이키바람막이
6 맨투맨	6 나이키바람막이	6 트위드자켓	6 트위드자켓
7 나이키후드티	7 트위드자켓	7 나이키바람막이	7 후드집업
8 나이키바람막이	8 후드집업	8 후드집업	8 후리스
9 트위드자켓	9 잠옷	9 후리스	9 잠옷
10 니트원피스	10 후드티	10 써스데이아일랜드원피스	10 바람막이

- 과거 책 단어 빈도
  - https://books.google.com/ngrams



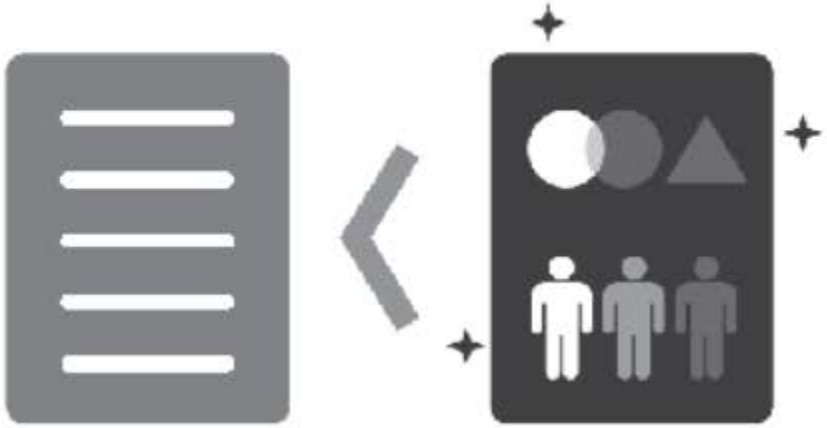
- 소셜분석체험
  - http://search.kbig.kr/experience/home.do

- 소셜분석체험
  - http://www.socialmetrics.co.kr/



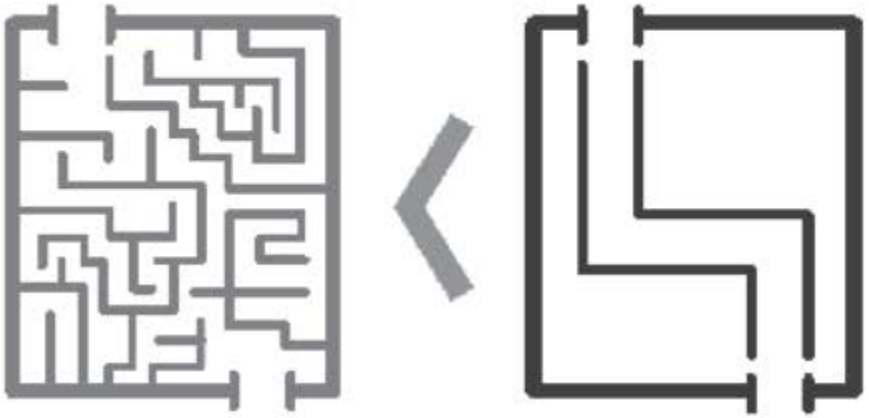
결국은 쌓아 놓은 데이터를 모든 분야에서 활용하는 시대
그 중심에 데이터 시각화가 각광

시각화는 직관적이고 시각적이기 때문에
소비자 혹은 독자에게 매력적인 콘텐츠 유형



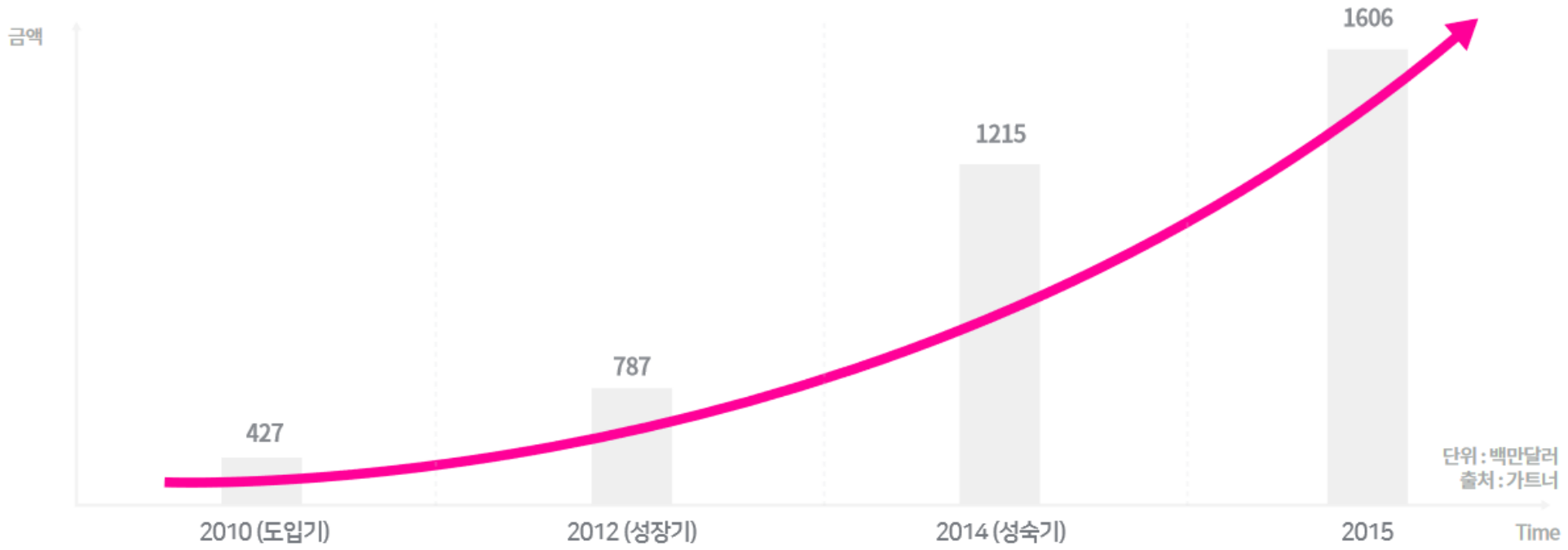
흥미 유발

이미지는 글보다 시각적인 즐거움을 줄 수 있다
딱딱하고 재미없는 정보를 부담스러워 하는
수용자들의 접근 장벽을 낮출 수 있다



이해 쉬운 전달

보통의 사람들이 당연히 알아야하고 필요한 정보들이
어렵고, 지루하게 존재하는 경우가 많다
이런 정보들은 쉬운 전달이 필요하다



특성

2010 (도입기)	2012 (성장기)	2014 (성숙기)	2015
데이터 시각화에 대한 관심 증가 개인 중심의 포트폴리오 작업 위주 시장	데이터 저널리즘과 정보시각화 교류 활발 Google chart와 같이 기본 차트형 프로그램 출현	국제기구, 해외 미디어, 대기업 등에서 시각화 콘텐츠 활용 활발 데이터 시각화 전문 미디어 등장 (NYT, VOX, FiveThirtyEight)	새로운 융합 분야 콘텐츠로 다양한 방면의 데이터와 디자인 융합 가능

국내

일러스트레이터를 활용한 인포그래픽(정보) 중심의 시각화

폭발하는 데이터를 활용한 데이터 시각화에 대한 관심 증대

들으면 잊는다

보면 기억한다

행동하면 이해한다

-공자-

시각화의 이해

- **데이터 시각화란**
  - 광범위하게 분산된 방대한 양의 자료를 분석해 한눈에 볼 수 있도록 도표나 차트 등으로 정리하는 것

- **얻을수 있는 효과**
  - 자료로부터 정보를 습득하는 시간 절감으로 즉각적인 상황 판단이 가능
  - 자료를 습득하는 사람들의 흥미를 유발하고 정보의 빠른 확산을 촉진
  - 자료를 기억하는데에도 도움

- **시각 이해의 위계**
  - 정보를 전달하여 이해시킨다는 측면에서 각각의 이해를 구축하고 있는 요소

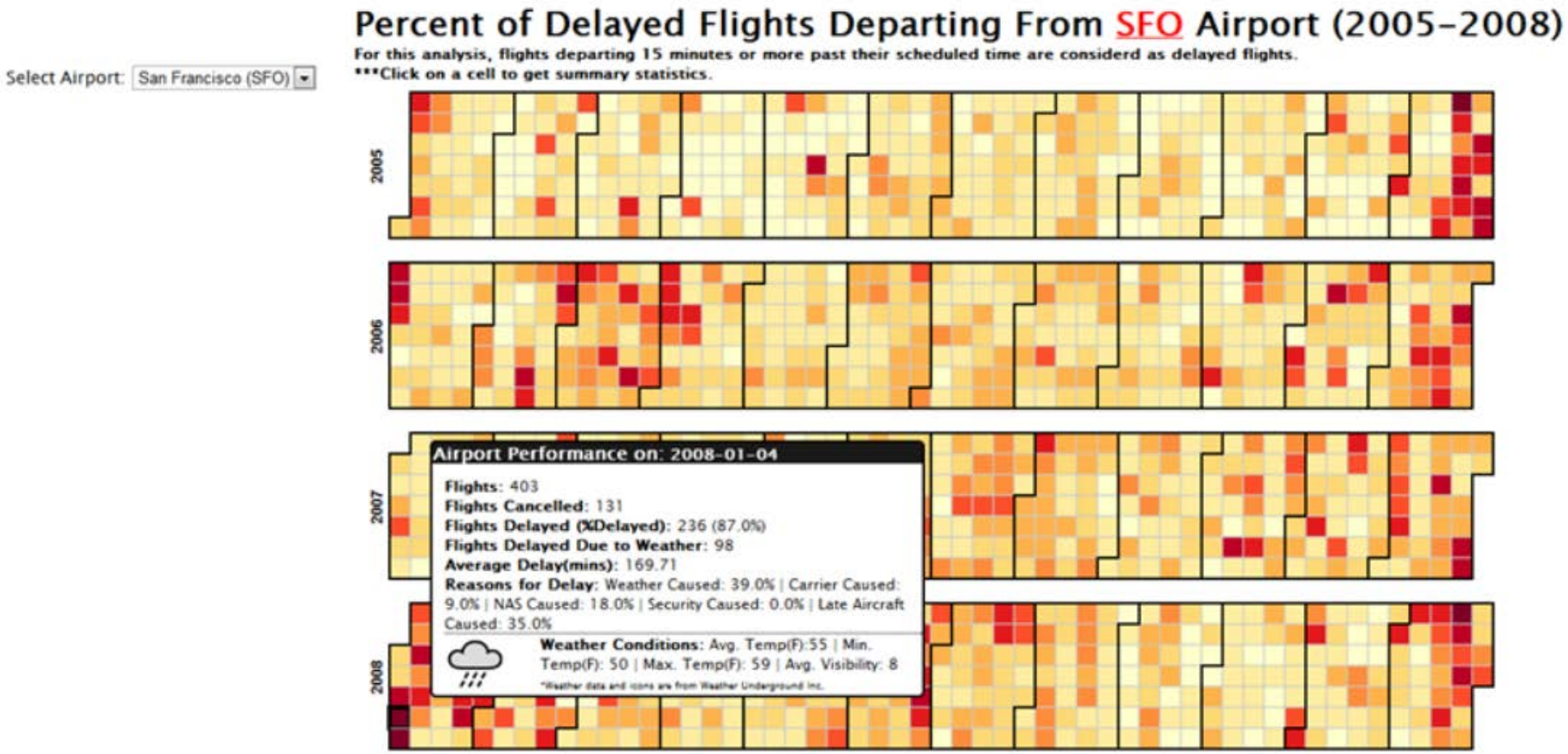
시각적 이해의 분류 체계는?

맥캔들레스가 그린 시각적 이해의 위계

http://www.informationisbeautiful.net/2010/data-information-knowledge-wisdom/

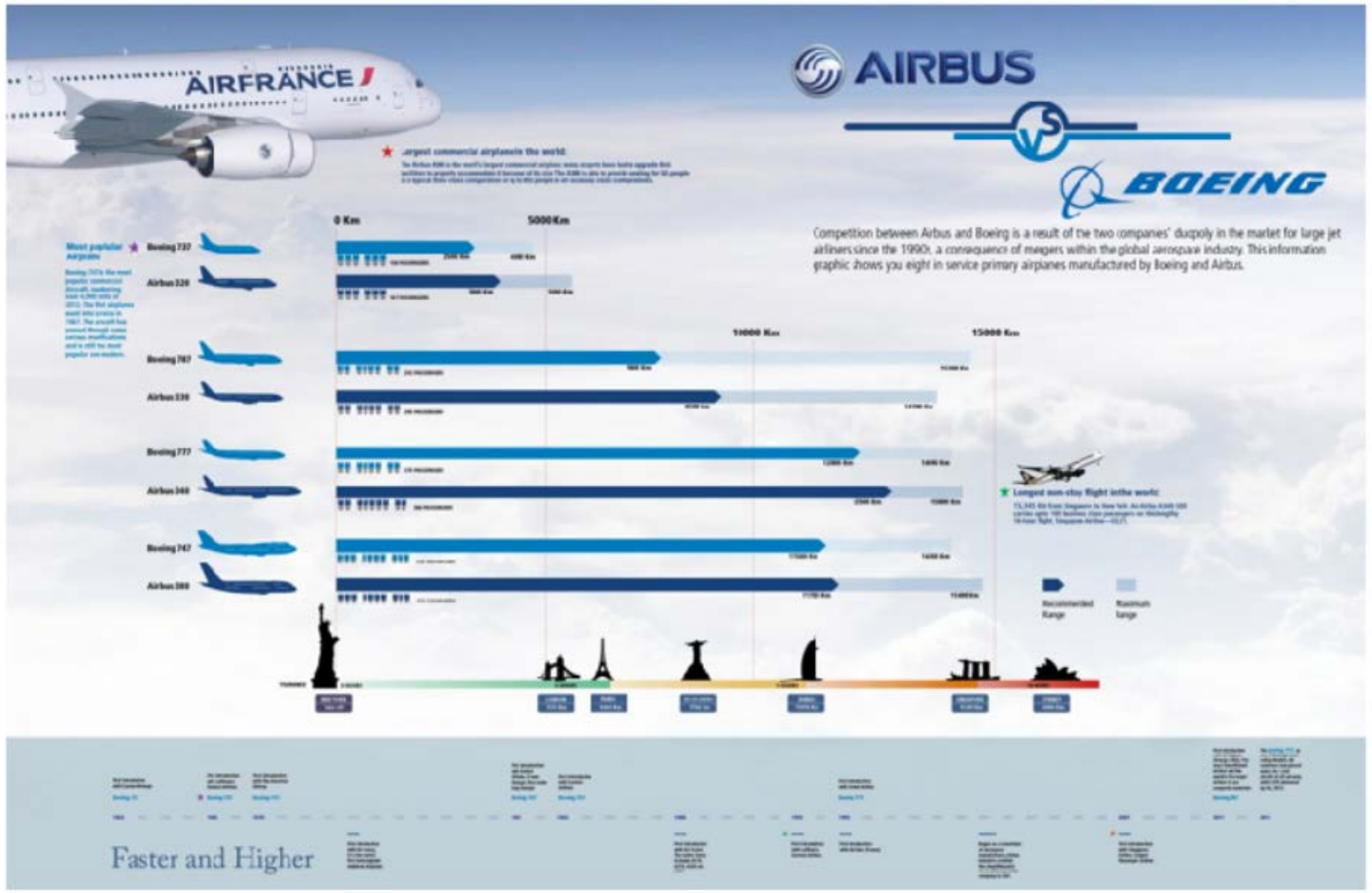
데 이 터 시 각 화

샌프란시스코 공항 출발 지연 일별 데이터 시각화



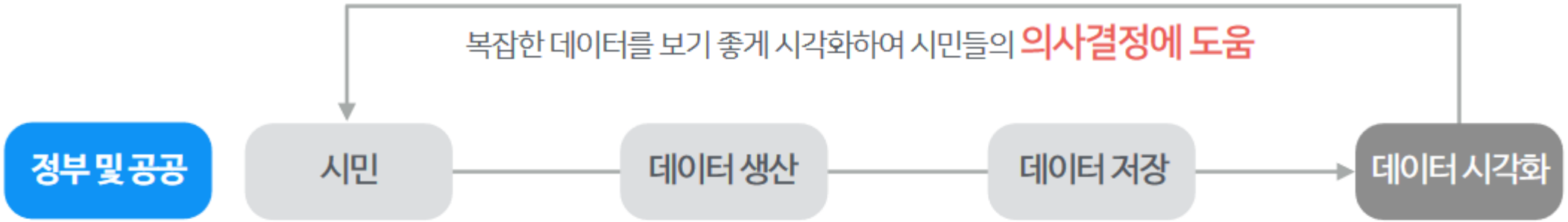
인 포 그 래 픽

Boeing vs. Airbus

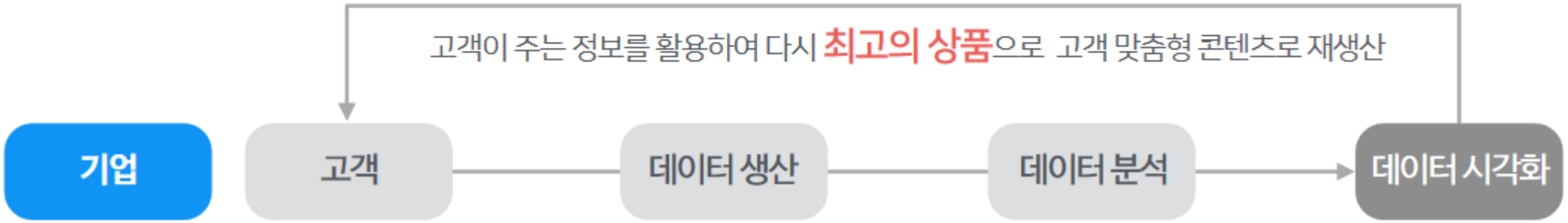


인포그래픽

- 중요한 정보를 한장의 그래픽으로 표현해 이를 보는 사람들이 손쉽게 해당 정보를 이해할 수 있도록 만드는 그래픽 메시지
- 패턴이나 트렌드를 보기 위해 인간의 시각능력에 집중
- 정보를 구체적, 표면적, 실용적으로 전달한다는 점에서 일반적인 그림이나 사진 등과는 구별되며, 복잡한 정보를 빠르고 명확하게 설명
- 뉴스 그래픽이라고 불리기도하며, 소셜 큐레이션의 수단으로 이용



"공공데이터 활성화? 핵심은 데이터 시각화"



국민권익위원회가 운영하는 민원빅데이터 개방시스템

한눈에 보는 **민원 빅데이터**

- 통계정보
- 분석정보
- 분석리포트
- 맞춤형 통계 및 사례

이용가이드 | 공공데이터 개방

오늘의 이슈

5	제차 신호조작 불이행	7
6	사건 재수사 및 처벌 촉구	7
7	Request Mungyeong South Ko administrative action	5
8	남한산성 터널 경사갱 및 환기탑 설치 반대	5

실시간 민원건수

16,748 건

전일 동시간 대비 ▲178.5%

민원 빅데이터 현황판의 통계 및 분석대상이 되는 민원은 기본적으로 국민신문고 민원과 지자체 민원을 포함합니다.

이용가이드 | 민원처리 사례

민원 건 수 현황

2020.10.05 16:13

시간대별 일별 월별 연도별

지역·기관별 현황

지역별 기관별

경기	서울	인천	경남	부산
6,459 건 ↑73.2%	2,251 건 ↑31.4%	868 건 ↑69.5%	635 건 ↑51.2%	601 건 ↑90.2%

오늘의 키워드

민원				언론				포털			
1	매교초 설립	6	김포시 고촌읍	1	코로나19	6	한국	1	현대 기아차 전기차 판매 세…	6	트럼프 의료진
2	불법 주정차	7	교육 환경	2	트럼프 대통령	7	백악관	2	코로나 확진 트럼프	7	과태료 10만원
3	주정차 신고	8	재산권 침해	3	미국	8	일본	3	상온노출 독감백신	8	한글날 집회
4	불법 주정차신고	9	신정차량기지이전 용역	4	중국	9	폼페이오	4	이르면 내일 퇴원	9	트럼프 측근
5	경기도 교육청	10	조합원 자격	5	코로나	10	서울	5	강원 산지 서리	10	노벨상 수상자 오늘부터 발표

민원키워드 클라우드

2020-09-06 ~ 2020-10-05

Copyright (c) 국민권익위원회 All Rights Reserved.

국민권익위원회

데이터 시각화

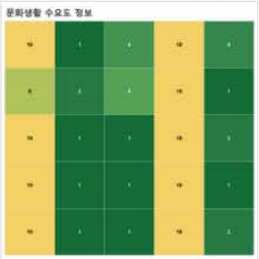
검색

아이돌 공식 유튜브 조회 트…

KStarLive 해외 방문자 분석

부산 행정동별 공급도 정보

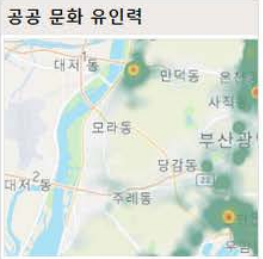
부산 타입별 공급도 지도


부산 행정동별 수요도 정보

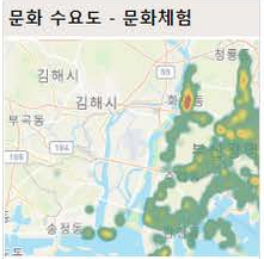
부산 타입별 수요도 지도

부산 문화시설 사각지대 정보

부산 문화소비 주 이용 지역 …



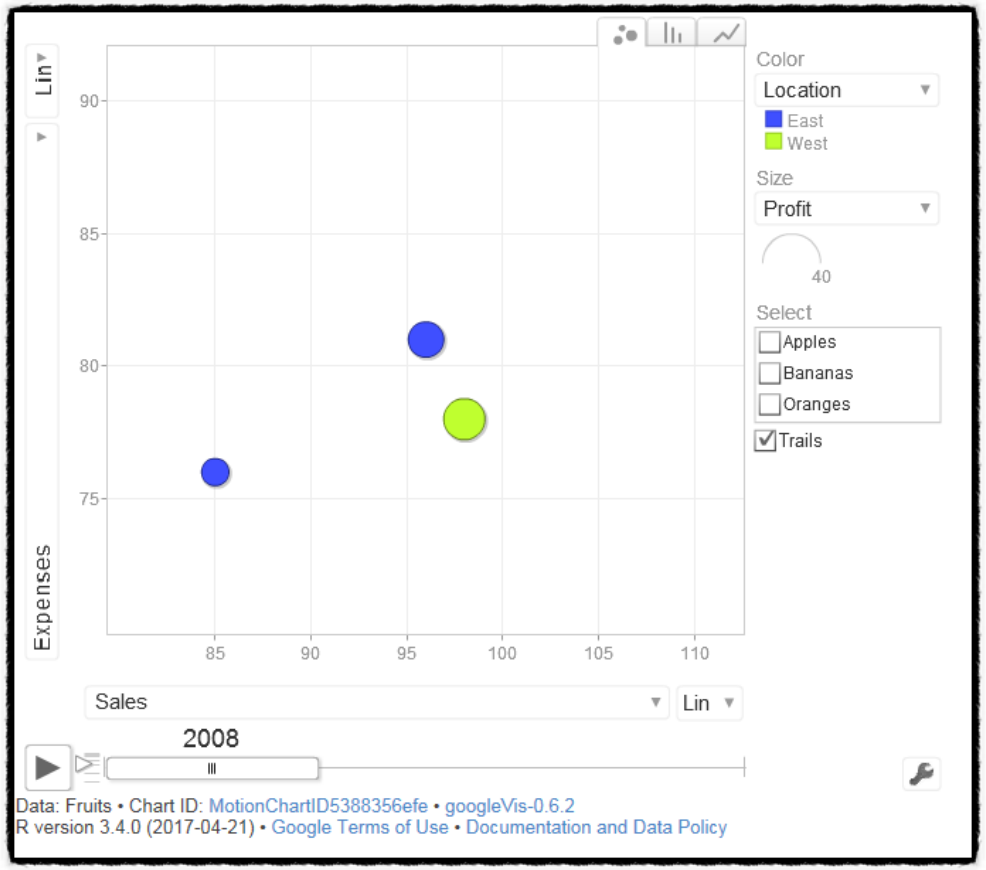
부산 공공/민간 유인력 정보



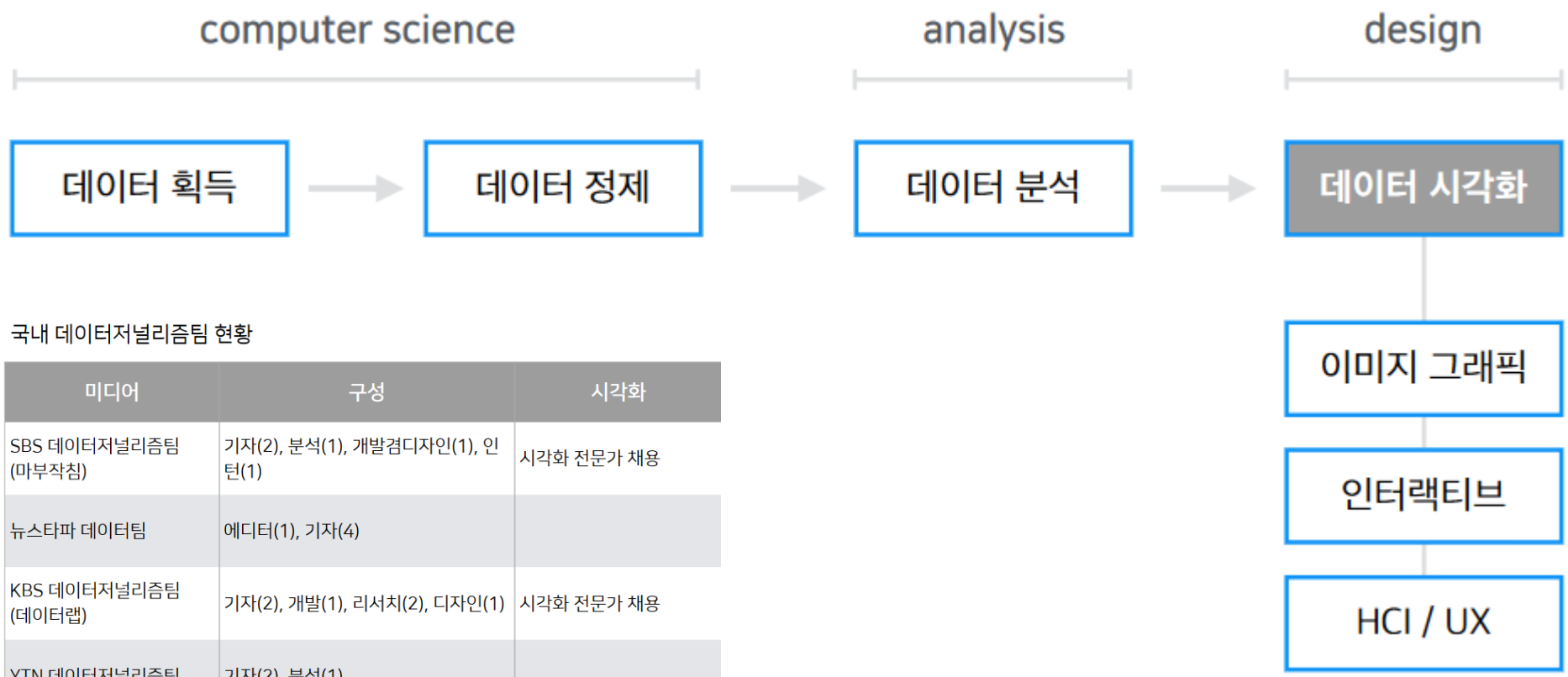
부산 문화생활 타입별 분포

R 시각화의 예

```r
if(!require(googleVis)) install.packages("googleVis")

library(googleVis)

Fruits

M1 <- gvisMotionChart(Fruits,
                      idvar="Fruit",
                      timevar="Year")

plot(M1)
```



시각화 방법론



국내 데이터저널리즘팀 현황

미디어	구성	시각화
SBS 데이터저널리즘팀 (마부작침)	기자(2), 분석(1), 개발겸디자인(1), 인턴(1)	시각화 전문가 채용
뉴스타파 데이터팀	에디터(1), 기자(4)	
KBS 데이터저널리즘팀 (데이터랩)	기자(2), 개발(1), 리서치(2), 디자인(1)	시각화 전문가 채용
YTN 데이터저널리즘팀	기자(2), 분석(1)	
한겨레 데이터	기자(1)	

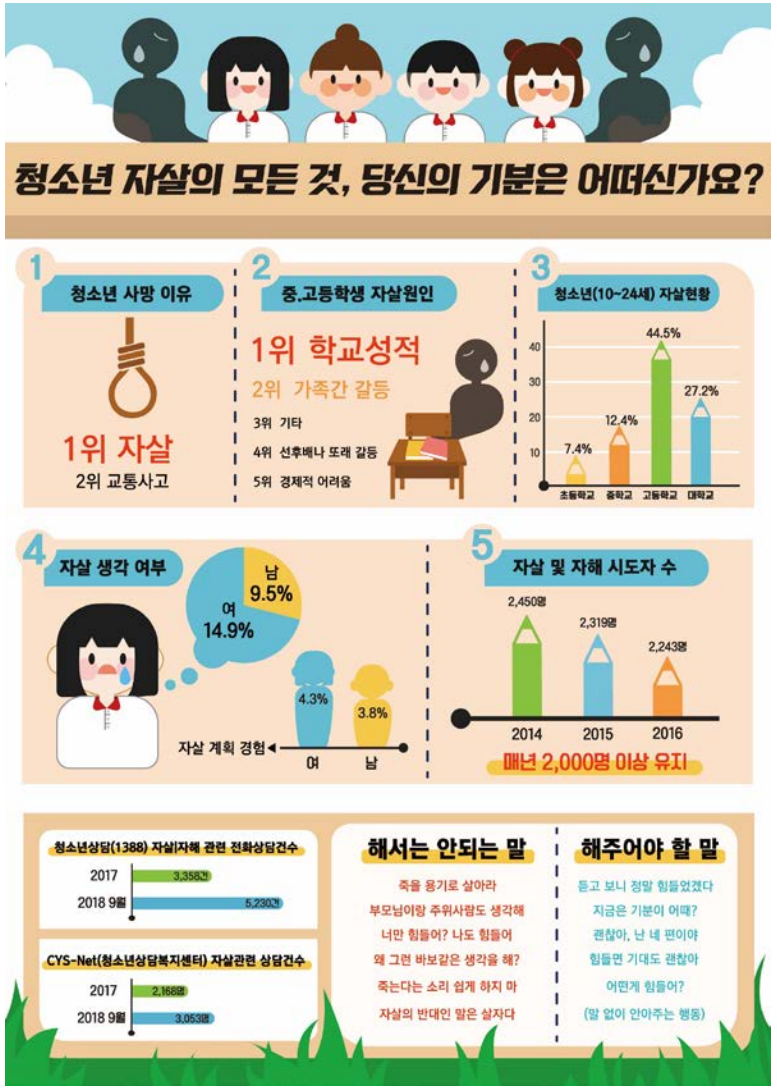
정보 디자인의 개념도

- 정보 디자인
  - 효과적인정보와 복잡하고 구조적이지 않은 기술데이터를 시각적으로 표현하는 방법
  - 인포그래픽 역시 정보 디자인의 한 유형임

인포그래픽 유형

- 통계기반
- 타임라인 기반
- 프로세스 기반
- 위치, 지리 기반
- 비교 형식
- 비주얼 스토리텔링

통계기반


청소년 자살의 모든 것, 당신의 기분은 어떠신가요?
1 청소년 사망 이유
1위 자살
2위 교통사고
2 중.고등학생 자살원인
1위 학교성적
2위 가족간 갈등
3위 기타
4위 선후배나 또래 갈등
5위 경제적 어려움
3 청소년(10~24세) 자살현황
7.4%
12.4%
44.5%
27.2%
초등학교 중학교 고등학교 대학교
4 자살 생각 여부
남 9.5%
여 14.9%
4.3%
3.8%
자살 계획 경험
여 남
5 자살 및 자해 시도자 수
2,450명
2,319명
2,243명
2014 2015 2016
매년 2,000명 이상 유지
청소년상담(1388) 자살|자해 관련 전화상담건수
2017 3,358건
2018 9월 5,230건
CYS-Net(청소년상담복지센터) 자살관련 상담건수
2017 2,168명
2018 9월 3,053명
해서는 안되는 말
죽을 용기로 살아라
부모님이랑 주위사람도 생각해
너만 힘들어? 나도 힘들어
왜 그런 바보같은 생각을 해?
죽는다는 소리 쉽게 하지 마
자살의 반대인 말은 살자다
해주어야 할 말
듣고 보니 정말 힘들었겠다
지금은 기분이 어때?
괜찮아, 난 네 편이야
힘들면 기대도 괜찮아
어떤게 힘들어?
(말 없이 안아주는 행동)

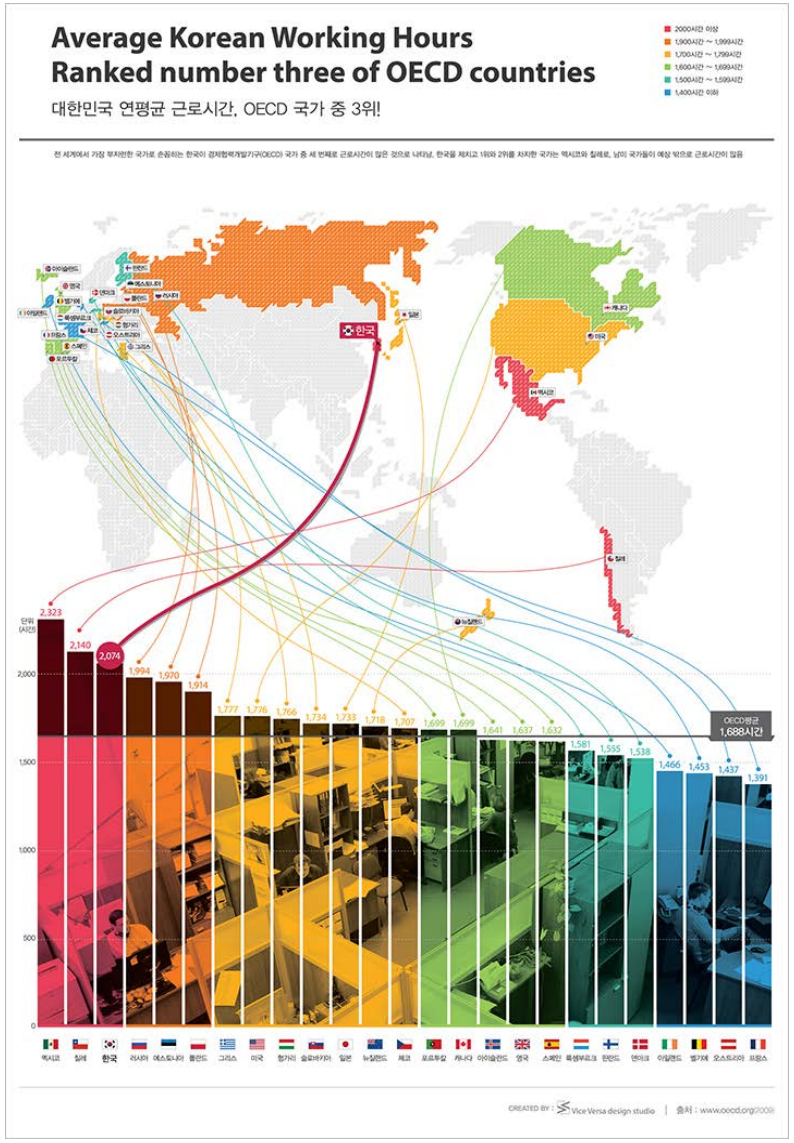
타임라인 + 통계

세명대
교환학생
되려면 무엇이 필요할까?
교환학생 가기에 좋은 시기
3학년1학기
3학년 2학기
선발방법
2차
1차
이메일 접수 및 오프라인 서류 제출
2017.09.21
학생준비
2017.10.08
면접 대상자 발표
2017.10.14
합격자 발표
나라별 대학 /선발총인원
2017.10.17
지원 대학 입학 서류 준비 및 발송
2017.10.28
합격자 오리엔테이션
2018.02.01
교환학생 파견
나라별 물가
말레이시아
일본
필리핀
한국

프로세스 기반

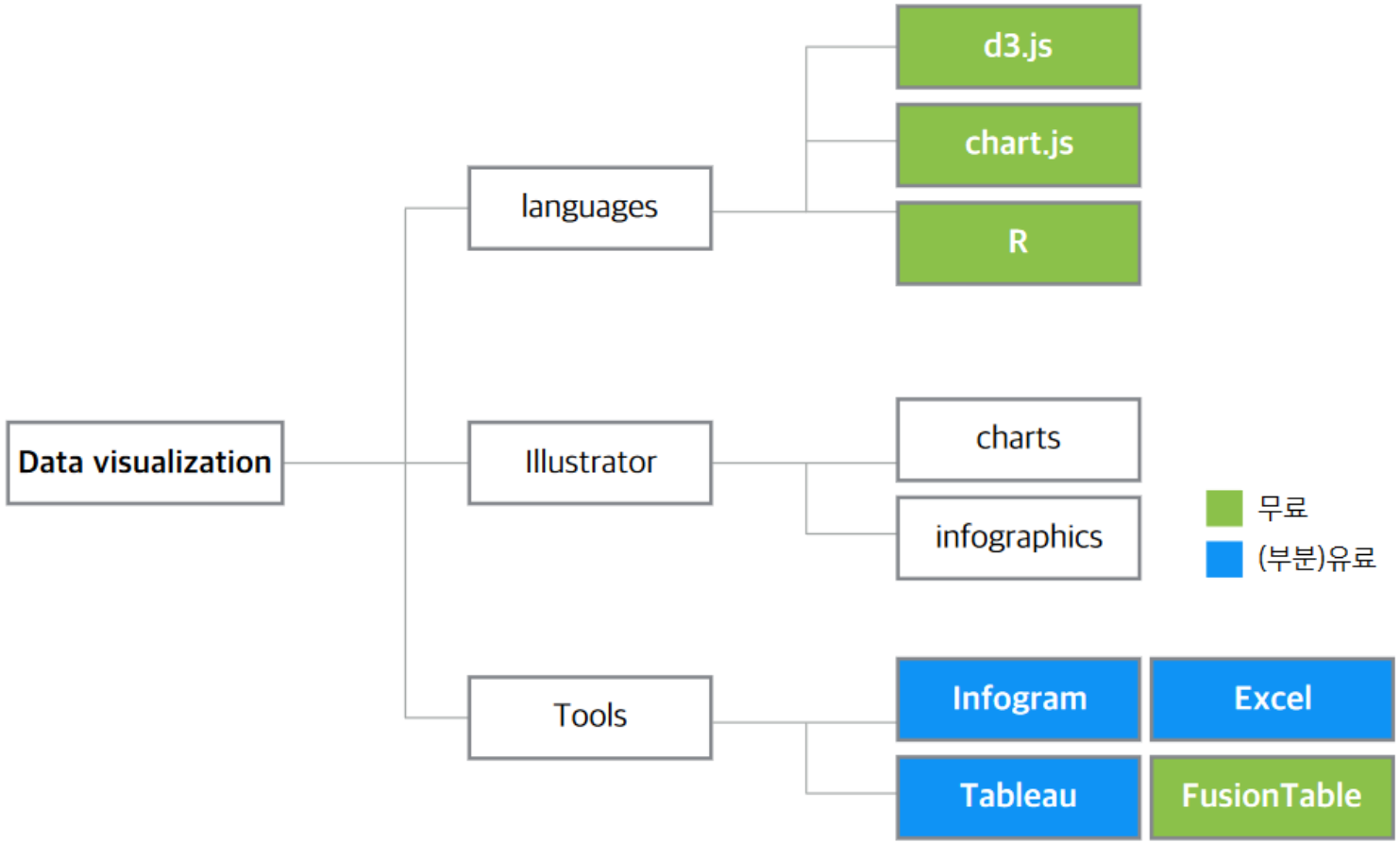
www.shutterstock.com · 734364490

위치,지리 기반



시각화 방법론

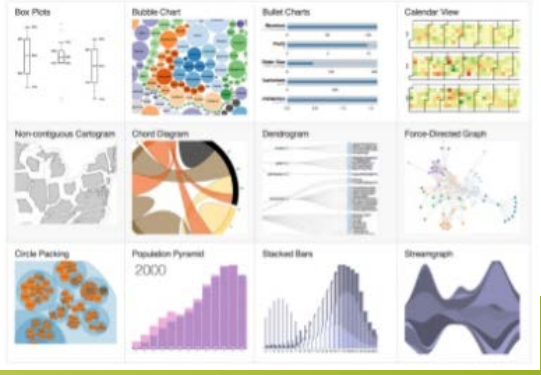
데이터시각화는 몇 가지 전문적인 언어와 툴을 활용



- 엑셀
  - 누구나 알고있고 쓰지만, 심도있는 분석이나 그래픽이 필요할 땐 쓰지 않음
  - 쉽게 그래프를 만들 수 있음, 11개 종류의 일반적인 차트를 그릴수 있음

- 구글 차트
  - 속도가 빠르며, 모바일에서도 잘 작동함
  - 동적인 데이터인 경우에는 구글 스프레드시트에 데이터를 저장한 후 Visualization Query로 읽어와 차트를 만들 수도 있음
  - 잘 모르는 초보자를 위한 위자드 툴도 제공
  - 스프레드쉬트와 차트 API를 함께 씀

- D3.js
  - 데이터 시각화 프레임워크, 자바스크립트로 개발
  - HTML, SVG, 그리고 CSS를 이용하여 시각화 요소를 만들어냄



다양한 인포그래픽 공모전

감사합니다